STANDARDS OF SOUTH AFRICAN SENIOR CERTIFICATE BIOLOGY EXAMINATIONS: 1994 TO 2007

VOLUME I: Chapters and References

Anna Aletta Crowe

Thesis presented for the degree of

DOCTOR OF PHILOSOPHY

in the School of Education

UNIVERSITY OF CAPE TOWN

July, 2012

Supervisors: Associate Professor Rüdiger Laugksch
Professor Timothy Dunne

For

Tim, Kim, Chris

and

Rachael and Mary − our future

DECLARATION

The thesis *Standards of South African Senior Certificate Biology Examinations: 1994 to 2007* contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge and belief, this thesis contains no material previously published by any other person except where due acknowledgement has been made.

Signature:

Anna Aletta Crowe

Date: 30 July 2012

PREFACE

My interest in assessment started as a parent − early conversations with my daughter about her school work taught me how important it was that educators be both explicit and articulate about assessments, in order to be fair to their students. Professionally, in my early career I practiced as a scientist and then I moved into science education − as a scientist with a social conscience, rather than as a social scientist. I have migrated between employment as an educator at secondary school level and at university level because of my particular interest in the interface between these two levels of South African education. These aspects of who I am, and my role since 1990 as an external moderator for a succession of South African bodies which certified the Senior Certificate (SC) Biology examinations each year, mean that a large portion of my life has been spent contemplating 'the meaning of assessments'. Over the years I have hoarded documents to do with SC Biology education, but it was only during this research that I started to understand the importance of having complete and accurate records about our educational history. This thesis is but one way, my way, of trying to make sense of the SC Biology examinations during a period of transition in South Africa education. In doing this work I have emerged with a greater appreciation for the complexity of the work that remains to be done and I have generated many more questions than those that I set out to answer at the start of this research. While I acknowledge the intellectual challenges and the political difficulties associated with explicating and understanding the standards associated with South African SC examinations, and indeed of the South African education system as a whole, we have to make progress in this area if we are to start to fulfil the educational promises, specifically about quality education for all, that were made to the people of South Africa in 1994.

ABSTRACT

Public examinations, such as the South African Senior Certificate (SC) examinations at the end of Grade 12, signal two messages to the society in which they operate: first, the competencies that are valued, that is, its standards; second, the required level of mastery in these competencies that are construed as indicators of success. The SC examinations certified successful students as competent to enter the workforce and, if they obtained a matriculation exemption, qualified them for admission to tertiary study. The SC was not a part of an explicit standards-based curriculum, and there is thus little understanding, but much public speculation, about the relationship between student achievement in the SC examinations, competency and standards.

In an attempt to understand this relationship—with a particular focus on the role of standards—in the SC Biology examinations over a period of time, the answer to the following research question was sought: *What did the SC Biology examinations in South Africa assess; did their focus change during the period 1994 to 2007; and, if so, what did this change mean?*.

Both in South Africa and internationally, ‘standards’ is an often-used educational term, the meaning of which has become confused in the literature and by public use. In this study, a methodology to make explicit the standards inherent within the SC Biology examinations—and the relationship between standards and student achievement—was developed, described and applied. The methodology involved using the international research literature about educational standards for the construction of a conceptual framework by which the standards of the South African SC Biology examinations could be extracted and understood. Validation processes give meaning to student performances in assessments, like the SC examinations. The conceptual framework developed in this study uses aspects of validity associated with the SC Biology examinations as a proxy for standards, and validity evidence is thus used to generate three strands of evidence necessary to impute standards from the examinations. The first two strands are the content standards generated from examination question papers and the performance standards from examination question papers and candidates’ answer scripts. The third strand concerns the structural aspects of the examination question papers known to influence student performance, and therefore the standards of an examination.

One element of content standards is cognitive demand. While, many different instruments have been used to determine the cognitive demand of tasks given to students, few of these have been empirically validated, and most lack any theoretical or empirical argument about how they function. Therefore, effective application of the conceptual framework for use in explicating the standards of SC Biology examinations required the development of an instrument, the

Performance Expectations Taxonomy (PET), to determine the cognitive demand of question papers. The PET is demonstrably both easy and reliable to use and is an amalgamation of a number of different instruments, used to classify the cognitive demand of assessments, including Bloom's Taxonomy (BTEO), combined with current knowledge about human cognition.

The standards were generated from analyses of the complete set of 111 SC Biology examination question papers, from all South African examining bodies in the important post-apartheid years from 1994 to 2007. A total of 11 006 scorable questions were generated from the analyses of question papers and 7 553 candidates' answer scripts were analysed for four examinations held in 2005 and 2006. Extracted standards varied between years and between examining bodies, and are described and compared in the light of the policies which directed the teaching and assessment of SC Biology and current global assessment practices. The varied content standards observed between the various SC Biology examinations analysed in this study brings into question the assumed equivalence of the examinations. The South African practice of certifying students in SC Biology examinations on the basis of one aggregate mark, sans meaning of that aggregate mark, together with the practice of using the same cut-scores each year to delineate performance standards in these examinations is shown to be flawed.

Findings from this study do not constitute judgement about the quality of the SC Biology examinations but employ a methodology developed specifically to understand standards in South African SC examinations from a defensible, research-based international perspective. The methodology permits evidence-based insights into the ways in which standards for a particular subject can vary across years and between examining authorities. The conceptual framework and its operationalization may be adapted for use in other SC examination subjects of previous years, but also for the current and planned examinations of the National Senior Certificate (NSC).

This research highlights the important and open question of how the South African education system must structure and gather validity evidence, an indicator of standards, to support, or challenge, the high-stakes inferences about students' knowledge that are made from examination performance at the end of their secondary schooling.

Keywords: Standards, South Africa, Senior Certificate, Biology, examinations, validity, cognitive demand.

LIST OF ABBREVIATIONS

Abbreviation	Full name
AAAS	American Association for the Advancement of Science
AERA	American Educational Research Association
ANC	African National Congress
APA	American Psychological Association
AQA	Assessment and Qualifications Alliance
BCVO	Beweging vir Christelik-Volkseie Onderwys
BOK	Breadth of Knowledge
BTEO	Bloom's Taxonomy of Educational Objectives
CASS	Continuous Assessment
CBS	Core Biology Syllabus
CCSSO	Council for Chief State School Officers
CED	Cape Education Department
COD	Code of Practice
DET	Department of Education and Training
DNE	Department of National Education
DoE	Department of Education
DOK	Depth of Knowledge
EALL	English as Additional Language Learner
ELL	English Language Learner
GCSE	General Certificate of Secondary Education
HG	Higher Grade
HOA	House of Assembly
HOCS	Higher Order Cognitive Skills
HOD	House of Delegates
HOR	House of Representatives
IEB	Independent Examinations Board
JMB	Joint Matriculation Board
KZN	Kwa-Zulu Natal
LG	Lower Grade
LOCS	Lower Order Cognitive Skills
MCQ	Multiple Choice Question
ME	Matriculation Endorsement
MENA	Middle East and North African
NAEP	National Assessment of Educational Progress

NAS	National Academy of Sciences
NCLB	No Child Left Behind
NCME	National Council on Measurement in Education
NCS	National Curriculum Statement
NDoE	National Department of Education
NEAP	National Assessment of Educational Progress
NEB	National Examinations Board
NED	Natal Education Department
NEPI	National Education Policy Investigation
NGDB	National Guideline Document for Biology
NRC	National Research Council
NSC	National Senior Certificate
OBE	Outcomes-based education
OECD	Organisation for Economic Co-operation and Development
OTL	Opportunity-to-learn
PET	Performance Expectations Taxonomy
PLD	Performance level description
PLL	performance level label
Safcert	South African Certification Council
SBA	School Based Assessment
SC	Senior Certificate
SG	Standard Grade
SI	Similarity Index
SIMS	Second International Mathematics Study
SoT	School of Tomorrow
TBVC	Transkei, Bophuthatswana, Venda, and Ciskei
TED	Transvaal Education Department
TED	Transvaal Education Department
TIMSS	The Third International Mathematics and Science Survey
TIMSS-R	TIMSS-Repeat
UK	United Kingdom
USA	United States of America
WCED	Western Cape Education Department
WCER	Wisconsin Center for Educational Research

LIST OF TABLES

		Page
Table 2.1	A summary of different education statistics by province (government schools only), 2008/2009	19
Table 2.2	SC HG and SG examination mark ranges and their equivalent symbols	23
Table 2.3	The point system used to rank SC marks by some South African universities for admission purposes	24
Table 2.4	Examining bodies responsible for the setting and administration of SC Biology examinations from 1994 to 2007	26
Table 2.5	Examples of changes in emphasis on different topics in SC during the period 1994 to 2007	31
Table 2.6	Summary of the sources of policy and documents which determined the version of the CBS used in 1994 to 2007	37
Table 2.7	Summary of the CBS Grades 10, 11and 12 according to the topics identified in the Augmented TIMSS Biology Curriculum Framework	43
Table 2.8	SC Biology Examination requirements for 1987 to 2007	45
Table 2.9	Number of students who wrote the government SC Biology HG and SG examinations 1994 to 2007	48
Table 2.10	Number of students who wrote the IEB SC Biology HG and SG examinations from 1994 to 2007	49
Table 2.11	A summary of SC examination results by province (government schools only), 2005 and 2006	55
Table 4.1	Percentage agreement between the PET classifications of individual raters and negotiated PET classifications of the group of five raters	151
Table 4.2	Objectives of the CBS used in this study (DNE, 1984a, 1984b) and their relationship to the categories of the PET	165
Table 4.3	Approach to the CBS used in this study (DNE, 1984a, 1984b) and their relationship to the major groups of the PET	166
Table 5.1	Total counts of WCED candidates and selected scripts analyzed for the 2005 SC Biology examinations	185
Table 5.2	Total counts of WCED candidates and selected scripts analyzed for the 2006 SC Biology examinations	188
Table 5.3	Invalid questions removed from script analyses	190
Table 5.4	A comparison on the percentage marks obtained by all candidates who wrote the 2005 and 2006 WCED SC Biology examinations, and by the samples of candidates selected for this study	191

Table 5.5	A list of topics examined in the Biology SC Examinations 1994 to 2007 as grouped in the CBS..	199
Table 6.1	Figures, tables and appendixes referred to in sections of this chapter	218
Table 6.2	Summary of marks allocated to different question types in 1994, 1995 CED and 1996-2000 WCED SC Biology HG question papers	223
Table 6.3	Summary of marks allocated to different question types in 1994-2000 IEB SC Biology HG question papers..	224
Table 6.4	Summary of marks allocated to different question types in 2001-2007 national DoE SC Biology HG and SG question papers...................................	225
Table 6.5	Summary of marks allocated to different question types in 2001-2007 IEB SC Biology HG and SG question papers...	226
Table 6.6	Types of questions, as a percentage of total marks, in the 1994, 1995 CED and 1996-2000 WCED SC Biology HG question papers.................................	235
Table 6.7	Types of questions, as a percentage of total marks, in the 2001-2007 national DoE SC Biology HG and SG question papers..	236
Table 6.8	Types of questions, as a percentage of total marks, in the 1994-2000 IEB SC Biology HG question papers...	237
Table 6.9	Types of questions, as a percentage of total marks, in the 2001-2007 IEB SC Biology HG and SG question papers...	238
Table 6.10	Essay topics for the 2001-2007 national DoE Biology HG question papers.....	247
Table 6.11	The rating scales used for the marking of essay-type questions in the 2001-2007 SC Biology examinations	248
Table 6.12	Essay topics for the 2001-2007 IEB SC Biology HG and SG question papers	249
Table 6.13	Types of free response answers, as a percentage of total marks, in the 1994, 1995 CED and 1996-2000 WCED SC Biology HG question papers...............	253
Table 6.14	Types of free response answers, as a percentage of total marks, in the 2001-2007 national DoE SC Biology HG and SG question papers...........................	254
Table 6.15	Types of free response answers, as a percentage of total marks, in the 1994-2000 IEB SC Biology HG question papers...	254
Table 6.16	Types of free response answers, as a percentage of total marks, in the 2001-2007 IEB SC Biology HG and SG question papers...	255
Table 6.17	Mean percentage weightings for topics in the 1994-2007 combined national DoE and IEB SC Biology HG and SG examinations....................................	258
Table 6.18	Percentage weightings of categories of cognitive demand in the 1994, 1995 CED and 1996-2000 WCED SC Biology HG question papers........................	267

Table 6.19 Percentage weightings of the different categories of cognitive demand in the 2001-2007 national DoE SC Biology HG and SG question papers.................. 267

Table 6.20 Percentage weightings of the different categories of cognitive demand in the 1994-2000 IEB SC Biology HG question papers... 268

Table 6.21 Percentage weightings of the different categories of cognitive demand in the 2001-2007 IEB SC Biology HG and SG question papers............................... 268

Table 6.22 Summary of the content similarity, measured by the similarity index (SI), 1994, 1995 CED,1996-2000 WCED and 2001-2007 national DoE Biology HG question papers... 287

Table 6.23 Summary of the content similarity, measured by the similarity index (SI), 2001-2007 national DoE Biology SG question papers.................................... 288

Table 6.24 Summary of the content similarity, measured by the similarity index (SI), 1994-2007 IEB Biology HG question papers... 289

Table 6.25 Summary of the content similarity, measured by the similarity index (SI), 2001-2007 IEB Biology SG question papers... 292

Table 6.26 Summary of the content range similarity, measured by the similarity index (SI), national DoE, 2001 to 2007 Biology HG and SG question papers.......... 293

Table 6.27 Summary of the content range similarity, measured by the similarity index (SI), 2001-2007 IEB Biology HG and SG question papers 294

Table 6.28 Summary of the content similarity, measured by the similarity index (SI), of Biology HG and SG question papers 295

Table 6.29 Correlations between the performances of candidates in the 2005 WCED examinations with respect to topic within each of the Papers 1 and 2 324

Table 6.30 Correlations between the performance of candidates in the 2006 WCED examinations with respect to topics within each of the Papers 1 and 2............ 325

Table 6.31 Correlations between the cognitive demand performance of candidates in the 2005 WCED examinations within each of the Papers 1 and 2 330

Table 6.32 Correlations between the cognitive demand performance of candidates in the 2006 WCED examinations with respect to cognitive demand within each of the Papers 1 and 2 331

Table 6.33 Correlations between the same candidates performance in each of the Papers 1 and 2 of the 2006 examinations with respect to cognitive demand 332

Table 6.34 Results of reliability analyses of groupings of scorable events: 2005 HG Paper 1.. 346

Table 6.35 Results of reliability analyses of groupings of scorable events: 2005 HG Paper 2 347

Table 6.36 Results of reliability analyses of groupings of scorable events: 2005 SG Paper 1.. 348

Table 6.37 Results of reliability analyses of groupings of scorable events: 2005 SG Paper 2 ... 349

Table 6.38 Results of reliability analyses of groupings of scorable events: 2006 HG Paper 1 ... 350

Table 6.39 Results of reliability analyses of groupings of scorable events: 2006 HG Paper 2 ... 351

Table 6.40 Results of reliability analyses of groupings of scorable events: 2006 SG Paper 1 ... 352

Table 6.41 Results of reliability analyses of groupings of scorable events: 2006 SG Paper 2 ... 353

Table 6.42 Results of factor analyses of 2005 and 2006 script data as individual scorable events ... 355

Table 6.43 Results of factor analyses of 2005 and 2006 script data combined into topics and performance expectations ... 356

Table 6.44 Results of factor analyses of 2005 and 2006 script data combined into topics. 356

Table 6.45 Results of factor analyses of 2005 and 2006 script data combined into performance expectations ... 357

Table 6.46 Results of factor analyses of 2005 and 2006 script data combined into content categories ... 357

Table 6.47 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2005 HG Paper 1 ... 359

Table 6.48 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2005 HG Paper 2 ... 360

Table 6.49 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2005 SG Paper 1 ... 361

Table 6.50 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2005 SG Paper 2 ... 362

Table 6.51 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2006 HG Paper 1 ... 363

Table 6.52 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2006 HG Paper 2 ... 364

Table 6.53 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2006 SG Paper 1 ... 365

Table 6.54 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2006 SG Paper 2 ... 366

Table 6.55 Features of secondary school exit level Mathematics and Science examinations of eight countries illustrating comparative examination difficulty, including South Africa.. 372

Table 6.56 Correlations between marks obtained in each of the five questions comprising the 2005 SC Biology examinations.. 390

Table 6.57 Correlations between marks obtained in each of the five questions comprising the 2006 SC Biology examinations.. 391

LIST OF FIGURES

		Page
Figure 1.1	Structure of thesis showing the relationships between the chapters, objectives, sub-research questions and the research question.....................	14
Figure 2.1	The relationship between Chapter 2 and the research question stated in Chapter 1...	17
Figure 2.2	A comparison of HG and SG marks and symbols. Shaded areas indicate symbols which denote failure..	24
Figure 2.3	The Biology SC examination cycle for state schools.................................	29
Figure 2.4	The relationships between the objectives and approaches to the content as stipulated in the CBS..	38
Figure 2.5	Evolution of interpretations of the CBS HG entry for vitamins..................	40
Figure 3.1	The relationship between Chapter 3 and the previous chapters..................	59
Figure 3.2	Relationships among the dimensions of the strategic framework for the implementation and the maintenance of standards in the USA.................	71
Figure 3.3	The SIMS conceptual design of curriculum...	82
Figure 3.4	The relationship between curricula components showing alignment analyses and proposed alignment analyses conducted in the USA............	83
Figure 3.5	An overview of a conceptual framework for understanding SC Biology standards by examination analyses..	109
Figure 4.1	The relationship between Chapter 4 and the previous chapters..................	115
Figure 4.2	Weighting of cognitive demand in examinations as stipulated in two SC Biology policy documents..	119
Figure 4.3	Explanation of cognitive demand categories and requirements stipulated in different SC Biology policies, 2001 to 2007..	120
Figure 4.4	Cognitive demand categories and requirements stipulated in different NSC Life Sciences policies, 2008 to 2010...	121
Figure 4.5	Categories of cognitive demand recognized by various authors which are directly and indirectly referenced...	131
Figure 4.6	Categories of the PET and the kinds of performance expectations identified for each category	137
Figure 4.7	Key to classifying a scorable event according to the PET...........................	138
Figure 4.8	Relationships between the categories of the PET.......................................	142

Figure 4.9	An example of two questions classified according category Memorize of the PET, showing different levels of complexity within each question.......	152
Figure 4.10	Relationship between the Performance Expectations Taxonomy developed in this study, the original BTEO and adaptations of Bloom's Taxonomy used in Biology studies..	155
Figure 4.11	The PET cross-referenced with BTEO (Bloom et al., 1956)	159
Figure 4.12	Comparison of DoE classifications and PET classifications of cognitive demand	160
Figure 4.13	Examples of PET classification of essay questions..................................	163
Figure 4.14	Examples of questions/scorable events classified using the Memorize category of PET and explanations for their classification in this group.......	169
Figure 4.15	Examples of questions/scorable events classified using the Perform-Routine-Procedures category of PET and explanations for their classification in this group..	171
Figure 4.16	Examples of questions/scorable events classified using the Explain category of PET and explanations for their classification in this group......	173
Figure 4.17	Examples of questions/scorable events classified using the Analyze category of PET and explanations for their classification in this group......	175
Figure 4.18	Examples of questions/scorable events classified using the Apply category of PET and explanations for their classification in this group.......	177
Figure 5.1	The relationship between Chapter 5 and the previous chapters..................	179
Figure 5.2	Examples of calculations using the BOK index	202
Figure 5.3	Example of calculations using the Similarity Index (SI)............................	207
Figure 6.1	Relationship between Chapter 6 and the preceding chapters......................	215
Figure 6.2	Examples of HG multiple-choice questions and their associated marks within and between different examining bodies in 1994............................	241
Figure 6.3	Examples of HG multiple-choice questions and their associated marks within and between different examining bodies in 1996............................	242
Figure 6.4	Examples of HG and SG multiple-choice questions and their associated marks within and between different examining bodies in 2006..................	243
Figure 6.5	Questions classified in this study as outside of the CBS............................	261
Figure 6.6	Variation in the composition of content (raw scores) for 1994,1995 CED, 1996-2000 WCED and 2001-2007 national DoE Biology HG examinations..	271

Figure 6.7	Variation in the composition of content (raw scores) 1994-2007 IEB Biology HG examinations. Shaded cells indicate differences between different profiles within a year	276
Figure 6.8	Relationships between breadth of knowledge (BOK) and depth of knowledge (DOK), 1994 to 2007	299
Figure 6.9	Relationships between breadth of knowledge (BOK) and depth of knowledge (DOK), 1994 to 2007	301
Figure 6.10	Results of a cluster analysis of the profiles of 1994-2007 IEB SC Biology HG question papers	302
Figure 6.11	Results of a cluster analysis of the profiles of 2001-2007 national DoE and IEB SC Biology HG and SG question papers	303
Figure 6.12	Student performance within answer type at different achievement levels	307
Figure 6.13	Student performance by answer type within achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2	309
Figure 6.14	Student performance by question length within achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2	311
Figure 6.15	Student performance by non-text elements in questions within achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2	314
Figure 6.16	Student performance by the length of answers required within achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2	317
Figure 6.17	Student performance by answers which require non-text elements within achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2	319
Figure 6.18	Student performance by topics within achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2	321
Figure 6.19	Student performance by PET categories with achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2	327
Figure 6.20	Average percentage obtained in content categories within achievement levels, 2005 and 2006 HG Paper 1	334
Figure 6.21	Average percentage obtained in content categories within achievement levels, 2005 and 2006 HG Paper 2	336
Figure 6.22	Average percentage obtained in content categories within achievement levels, 2005 and 2006 SG Paper 1	338
Figure 6.23	Average percentage obtained in content categories within achievement levels, 2005 and 2006 SG Paper 2	340
Figure 6.24	Total possible marks for content categories – Paper 1	342

Figure 6.25	Total possible marks for content categories – Paper 2................................	343
Figure 6.26	Relationship between symbols achieved in HG Paper 1, HG Paper 2 and the aggregate mark used for certification, 2006 sample scripts only...........	368
Figure 6.27	Relationship between symbols achieved in SG Paper 1, SG Paper 2 and the aggregate mark used for certification, 2006 sample scripts only...........	369
Figure 6.28	Mean mark per scorable event for 2005 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by question type	375
Figure 6.29	Mean mark per scorable event for 2006 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by question type	376
Figure 6.30	Mean mark per scorable event for 2006 HG Paper 1 and Paper 2 and SG Paper 1and Paper 2, by answer type..	377
Figure 6.31	Mean mark per scorable event for 2006 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by answer type..	378
Figure 6.32	Mean mark per scorable event for 2005 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2 by topic..	379
Figure 6.33	Mean mark per scorable event for 2006 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by topic..	380
Figure 6.34	Mean mark per scorable event for 2005 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by PET category..	382
Figure 6.35	Mean mark per scorable event for DoE 2006 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by PET category..	383
Figure 6.36	Mean mark per scorable event for DoE 2006 HG Paper 2 and DoE 2006 SG Paper 2 by BTEO categories	384
Figure 6.37	WCED HG and SG student performance according to aggregate marks, 2005 to 2007...	387
Figure 6.38	Examination inputs and information about students' abilities....................	393
Figure 6.39	Relationships between the different sections of this chapter and Chapter 7	396
Figure 7.1	The structure of this enquiry into SC Biology examinations and student performances	399

LIST OF APPENDICES

		Page
Appendix 2.1	An example of the principles of marking attached to each marking memorandum (DoE, 2004d)..	485
Appendix 4.1	Summaries of other taxonomies of cognitive demand considered, and rejected, for use in this study ………………………………….…………	487
Appendix 4.2	Table used by raters to code the performance expectations of a scorable event during validation of the PET……………………….......................	491
Appendix 4.3	The key used by raters to code the performance expectations of a scorable event during validation of the PET..	492
Appendix 4.4	Instructions for raters during validation of the PET....................................	493
Appendix 4.5	Table used by raters to code the topics of a scorable event during validation of the PET..	494
Appendix 4.6	Handout used to orientate raters involved in validation of the PET............	495
Appendix 4.7	Data capture sheets for the analysis of 2005 Biology Higher Grade Paper 1 − validation of PET..	497
Appendix 4.8	Data capture sheets for the analysis of 2005 Biology Higher Grade Paper 2 − validation of PET	499
Appendix 4.9	Table used by raters to check classifications during the PET post-validation. Shaded cells were contested and negotiated.............................	501
Appendix 5.1	List of SC Biology examination question papers analyzed in this study.....	507
Appendix 5.2	The coding form used to extract data from examination question papers..	509
Appendix 5.3	The number of scorable events recognized in the biology questions papers analyzed in this study...	511
Appendix 5.4	The coding form used to compile the profiles of examination question papers..	513
Appendix 5.5	Summaries of the questions used to compile the profile(s) for each of the SC Biology question papers analyzed in this study.............................	519
Appendix 6.1	Summary of marks allocated to different types of questions for 1994 Biology HG question papers..	528
Appendix 6.2	Summary of marks allocated to different types of questions for 1995 Biology HG question papers..	529
Appendix 6.3	Summary of marks allocated to different types of questions for 1996 Biology HG question papers..	530

Appendix 6.4 Summary of marks allocated to different types of questions for 1997 Biology HG question papers........ 531

Appendix 6.5 Summary of marks allocated to different types of questions for 1998 Biology HG question papers........ 532

Appendix 6.6 Summary of marks allocated to different types of questions for 1999 Biology HG question papers........ 533

Appendix 6.7 Summary of marks allocated to different types of questions for 2000 Biology HG question papers........ 534

Appendix 6.8 Summary of marks allocated to different types of questions for 2001 to 2007 national DoE Biology HG question papers........ 535

Appendix 6.9 Summary of marks allocated to different types of questions for 2001 to 2007 national DoE Biology SG question papers........ 536

Appendix 6.10 Summary of marks allocated to different types of questions for 2001 to 2007 IEB Biology HG and SG question papers........ 537

Appendix 6.11 Types of questions, as percentages of total marks, in the SC Biology HG question papers 1994........ 538

Appendix 6.12 Types of questions, as percentages of total marks, in the SC Biology HG question papers 1995........ 539

Appendix 6.13 Types of questions, as percentages of total marks, in the SC Biology HG question papers 1996........ 540

Appendix 6.14 Types of questions, as percentages of total marks, in the SC Biology HG question papers 1997........ 541

Appendix 6.15 Types of questions, as percentages of total marks, in the SC Biology HG question papers 1998........ 542

Appendix 6.16 Types of questions, as percentages of total marks, in the SC Biology HG question papers 1999........ 543

Appendix 6.17 Types of questions, as percentages of total marks, in the SC Biology HG question papers 2000........ 544

Appendix 6.18 Types of questions, as percentages of total marks in the national DoE and IEB SC Biology HG question papers 2001-2007........ 545

Appendix 6.19 Types of questions, as percentages of total marks in the national DoE and IEB SC Biology SG question papers 2001-2007........ 546

Appendix 6.20 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1994........ 547

Appendix 6.21 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1995........ 547

Appendix 6.22 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1996.. 548

Appendix 6.23 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1997.. 549

Appendix 6.24 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1998.. 550

Appendix 6.25 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1999.. 551

Appendix 6.26 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 2000.. 552

Appendix 6.27 Types of free response answers, as percentages of total marks in the national DoE and IEB SC Biology HG question papers 2001-2007........... 553

Appendix 6.28 Types of free response answers, as percentages of total marks in the national DoE and IEB SC Biology SG question papers 2001-2007............ 554

Appendix 6.29 Percentages per topic in the 1994 SC Biology HG examinations............... 555

Appendix 6.30 Percentages per topic in the 1995 SC Biology HG examinations............... 556

Appendix 6.31 Percentages per topic in the 1996 SC Biology HG examinations............... 557

Appendix 6.32 Percentages per topic in the 1997 SC Biology HG examinations............... 558

Appendix 6.33 Percentages per topic in the 1998 SC Biology HG examinations.............. 559

Appendix 6.34 Percentages per topic in the 1999 SC Biology HG examinations............... 560

Appendix 6.35 Percentages per topic in the 2000 SC Biology HG examinations.............. 561

Appendix 6.36 Percentages per topic in the2001-2007 national DoE and IEB SC Biology HG examinations.. 562

Appendix 6.37 Percentages per topic in the 2001-2007nnational DoE and IEB SC Biology SG examinations.. 563

Appendix 6.38 Percentage weightings of PET categories of cognitive demand in the 1994 SC Biology HG question papers.. 565

Appendix 6.39 Percentage weightings of PET categories of cognitive demand in the 1995 SC Biology HG question papers.. 565

Appendix 6.40 Percentage weightings of PET categories of cognitive demand in the 1996 SC Biology HG question papers... 566

Appendix 6.41 Percentage weightings of PET categories of cognitive demand in the 1997 SC Biology HG question papers... 566

Appendix 6.42 Percentage weightings of PET categories of cognitive demand in the 1998 SC Biology HG question papers... 567

Appendix 6.43 Percentage weightings of PET categories of cognitive demand in the 1999 SC Biology HG question papers.. 567

Appendix 6.44 Percentage weightings of PET categories of cognitive demand in the 2000 SC Biology HG question papers.. 568

Appendix 6.45 Percentage weightings of categories of PET cognitive demand in the national DoE and IEB SC Biology HG question papers, 2001-2007........... 568

Appendix 6.46 Percentage weightings of PET categories of cognitive demand in the national DoE and IEB SC Biology SG question papers, 2001-2007........... . 569

Appendix 6.47 Relationship between BOK and DOK in the 1994 SC Biology HG question papers.. 570

Appendix 6.48 Relationship between BOK and DOK in the 1995 SC Biology HG question papers.. 571

Appendix 6.49 Relationship between BOK and DOK in the 1996 SC Biology HG question papers.. 572

Appendix 6.50 Relationship between BOK and DOK in the 1997 SC Biology HG question papers.. 573

Appendix 6.51 Relationship between BOK and DOK in the 1998 SC Biology HG question papers.. 574

Appendix 6.52 Relationship between BOK and DOK in the 1999 SC Biology HG question papers.. 575

Appendix 6.53 Relationship between BOK and DOK in the 2000 SC Biology HG question papers.. 576

Appendix 6.54 Relationship between BOK and DOK in the 2001-2007 national DoE and IEB SC Biology HG question papers 577

Appendix 6.55 Relationship between BOK and DOK in the 2001-2007 national DoE and IEB SC Biology SG question papers.. 578

Appendix 6.56 Cluster analysis of the profiles of SC Biology HG question papers analyzed in this study.. 579

Appendix 6.57 Content categories 2005 and 2006 HG and SG Paper 1 – Summary of the performance level at which content categories become absent from students scripts.. 580

Appendix 6.58 Content categories 2005 and 2006 HG and SG Paper 2 – Summary of the performance level at which content categories become absent from the analysis of students scripts... 581

Appendix 6.59 Questions with mean percentage achieved greater than 80% or less than 20% in the 2005 national DoE SC HG and SG Biology examinations and PET category.. 582

Appendix 6.60 Questions with mean percentage achieved greater than 80% or less than 20% in the 2006 national DoE SC HG and SG Biology examinations and PET category.. 584

Appendix 7.1 Series of possible research questions generated by this study.................... 585

ACKNOWLEDGEMENTS

I am privileged to have had the opportunity and the space to think about my own, and South African, assessment practices and to discover what the rest of the world is doing with respect to assessment. However, in doing this thesis I have learned so much more about education than just about assessment. I know that my supervisors, Associate Professor Rüdiger Laugksch and Professor Timothy Dunne, often considered unnecessary my many ventures outside of the problem as they saw it, but they allowed me the freedom to go there (and sometimes had little option but to travel with me) to satisfy my curiosity and quest to know just that little bit more. For their enthusiasm, and the intellectual and emotional support that both these gentlemen gave me throughout my study, and for their generosity of ideas and advice, I will always be most grateful. There were so many times when I thought I was clearly communicating what was going on in my head, and I was not. From Rudi I learned the importance of clarity of thought, of conceptual frameworks in research, and to use commas responsibly in my writing. Tim taught me the minutia of always being explicit, that it is okay to challenge the status quo, and he tried to ensure that I was never over-influenced by a social science perspective of assessment. Thank you Rudi and Tim, for your time, your patience and your wisdom.

Through numerous conversations, some in corridors, with Associate Professor Rob Siebörger I came to better understand how connected and universal knowledge is, and that scientific ways of knowing are not really that different to historical ways of knowing. Thank you Rob, for your willingness to engage with my endless questions and ideas.

Three months into my research Professor Harold Noah commented on the first draft of my research proposal at a Spencer Foundation meeting of doctoral students. I am grateful for his enthusiastic, and more importantly, kind comments about what were at that stage my very early, and somewhat naïve, attempts to understand educational standards.

Professors Robert Brennan, David Conley, Ed Haertel, Bob Lissitz and Andrew Porter, took time from their busy AERA Annual Meeting schedules to discuss my research with me. Their encouragement was invaluable. Professors Lissitz and Saul Geiser willingly confirmed my interpretations and adaptations of certain measurement literature.

Former Umalusi colleagues, Professor Bill Fraser and Dr Dev Isaac, are thanked for numerous discussions about Senior Certificate examinations while we worked, and for their continued encouragement and interest in my research.

The Western Cape Education Department (WCED), through Dr Ronald Cornelissen, gave permission for me to use samples of SC Biology examination scripts from their SC Biology candidates. Errol van Breda and staff of the WCED Examinations Section carefully batched and labelled the sample scripts for me. Special thanks must go to the unknown 2005 and 2006 Grade 12 students whose SC Biology scripts formed the basis of the performance analyses presented in this thesis.

Dr Rufus Poliah and Priscilla Ogunbanjo of the Department of Basic Education (formerly the Department of Education); Dr Rakometsi, Emmanuel Sibanda, and Paul Mokilane of Umalusi; Cobus Lötter of the Matriculation Board; Anne Oberholzer, Estelle Nel, Victoria Masasanya and Jenny Page of the Independent Examinations Board; Tommy Botha, Theo Hamman, David Green and Fadli Wagiet of the WCED; and Peter Ayerst, all willingly helped me to find disparate bits of information that I needed to complete my story.

Associate Professor Rudi Laugksch, Drs Dev Isaac and Melanie Luckay, and Tommy Botha helped me to refine the instrument developed to measure cognitive demand in this study. Dr Alison Crowe (no relation) allowed me access to her classifications of the cognitive demand of university biology examination questions using her adapted version of Bloom's Taxonomy of Educational Objectives.

Staff of the University of Cape Town Libraries, especially those from the Interlibrary Loans Department, helped me to locate many reference articles. Staff of the Educational Testing Service (ETS) and Hilary Nicolls of the Assessment and Qualifications Alliance (AQA); Drs Michael Booker, Robert Brennan, William Crowe, Sam Geiser, Maria Martiniello, Curtis McKnight, Martie Sanders, Stephen Sireci, Ken Travers and Gilbert Valverde kindly sent me copies of articles I had difficulty obtaining locally.

Dr Graeme Oatley, and Vivienne Katshunga, Charles Mbayi, Jean-Marie Sabwa, Mbane Thuotou Sayi, joined me in computerizing the question paper and script data. Carrie-Anne Urquhart helped me to edit the figures, tables and appendices of Chapter 6. Thank you all for your help and the accuracy with which you did these jobs.

Staff of the School of Education, especially Ingrid Thom, Chris Kleinsmith and Lance MacLeod made my time there both enjoyable and comfortable. My PhD colleagues, especially Drs Melanie Luckay, Caroline Long, Leadus Madzima and Robyn Sokalinski made a shared journey so much easier. Caroline kindly commented on Chapter 1 of this dissertation and was so positive about my research, especially during the last phases of my writing. My niece, Charmaine Urquhart, and my friend, Colleen Callahan, encouraged me throughout my study and regularly reminded me of the challenges faced on a daily basis by South African classroom teachers. Thank you all for your support.

I owe enormous gratitude to all my family, friends, colleagues and students who have for so many years indulged and 'fed' my passion for teaching and learning, and accepted my uncompromising insistence on equity in education.

My most important acknowledgement belongs to my family. My husband Timothy Crowe enabled me to be a 'kept' woman, and indeed encouraged me, to pursue my ideas. Tim was my private reprographic assistant and engaged in numerous (heated) debates with me about the meaning of validity. From the other side of the world my daughter Kimberley Crowe Larsen and her husband Christian Larsen helped me to solve many computer mysteries, and also gifted me with books that I had coveted but was unable to purchase easily, or affordably, in South Africa. Thank you Tim, Kim and Chris for your unconditional love, patience and support, and most importantly, for believing in me.

This work was supported in part by the Spencer Foundation, the Andrew Mellon Foundation and the National Research Foundation, and is hereby gratefully acknowledged.

TABLE OF CONTENTS

Page

VOLUME I

DEDICATION ………………………………………………………… i

DECLARATION ………………………………………………………… ii

PREFACE ………………………………………………………… iii

ABSTRACT ………………………………………………………… iv

LIST OF ABBREVIATIONS ………………………………………………………… vi

LIST OF TABLES ………………………………………………………… viii

LIST OF FIGURES ………………………………………………………… xiii

LIST OF APPENDICES ………………………………………………………… xvii

ACKNOWLEDGEMENTS ………………………………………………………… xxii

TABLE OF CONTENTS ………………………………………………………… xxv

CHAPTER 1 **INTRODUCTION** ………………………………………………………… 1

1.1 **Background and rationale** ………………………………………………………… 2

1.2 **Problem statement** ………………………………………………………… 7

1.3 **Aim** ………………………………………………………… 7

1.4 **Objectives**………………………………………………………… 7

1.5 **Research question and sub-questions** ………………………………………………………… 8

1.5.1 Research question ………………………………………………………… 8

1.5.2 Research sub-questions ………………………………………………………… 9

1.6 **Clarification of terms** ………………………………………………………… 10

1.7 **Organization of the remainder of this thesis** ………………………………………………………… 12

CHAPTER 2 **CONTEXTUAL FACTORS** ………………………………………………………… 17

2.1 Processes, events and procedures associated with the examination of Senior Certificate Biology in South Africa (1994 to 2007) ………………………………………………………… 20

2.1.1 The South African Senior Certificate ………………………………………………………… 21

2.1.2 The South African Senior Certificate examining bodies ………………………………………………………… 25

2.1.3 The South African Senior Certificate examination cycle ………………………………………………………… 27

2.1.3.1 *Setting of question papers* ………………………………………………………… 28

2.1.3.2 *Marking review* . ………………………………………………………… 33

2.1.3.3 *Standardization of marks*………………………………………………………… 33

2.1.4 The South African Senior Certificate Biology Syllabus 36
2.1.5 The South African Senior Certificate Biology examination requirements 42
2.2 **Areas of research focus** 46
2.2.1 Why focus on Biology? 46
2.2.2 Why focus on Senior Certificate? 50
2.2.3 Why select the period 1994 to 2007? 51
2.2.4 Why use Western Cape Education Department candidates' answer scripts? 54
2.3 **Comparative international studies of science assessment systems including examinations** 55
2.4 **Chapter summary** 58

CHAPTER 3 **DEVELOPMENT OF A CONCEPTUAL FRAMEWORK FOR DESCRIBING AND ANALYZING STANDARDS OF SOUTH AFRICAN SENIOR CERTIFICATE BIOLOGY EXAMINATIONS** 59
3.1 **Standards in education** 62
3.1.1 What are educational standards? 62
3.1.2 Standards in the United States of America 68
3.1.3 Standards in the United Kingdom, Australia and the Netherlands 74
3.1.4 Standards in South Africa 78
3.1.5 What it means to have a standards-based curriculum 80
3.2 **Validity and assessment** 90
3.2.1 The evolution of the concept validity as applied to assessment 91
3.2.2 Sources of validity evidence 98
3.3 **Alignment** 102
3.4 **Equating of assessments** 104
3.5 **A conceptual framework to explicate standards from South African Senior Certificate Biology examinations** 108
3.5.1 A description of the conceptual framework 108
3.5.2 The significance of the conceptual framework for this study 111
3.5.3 Working definitions of standards used in the framework and from here on in this thesis 112
3.6 **Chapter summary** 113

CHAPTER 4 **DEVELOPMENT OF AN INSTRUMENT TO DETERMINE COGNITIVE DEMAND** 115
4.1 **Policy with respect to the cognitive demand of Biology Senior Certificate examinations** 117
4.2 **Measures of cognitive demand** 122

4.2.1 The quest for a way to measure cognitive demand 123
4.2.2 Bloom's Taxonomy of Educational Objectives.......... 126
4.2.3 An example of misinterpretation and misrepresentation of Bloom's Taxonomy of Educational Objectives 130
4.3 **The Performance Expectations Taxonomy** 134
4.3.1 Description of the Performance Expectations Taxonomy.......... 135
4.3.1.1 *Categories of performance expectations* 136
4.3.1.2 *Relationships between the categories of performance expectations* 142
4.3.1.3 *Different ways of grouping categories of performance expectation* 143
a ***Rote-And-Routine vs Demonstrate-Understanding*** 144
b ***Lower order cognitive skills vs higher order cognitive skills*** 147
4.3.2 Validation of the Performance Expectations Taxonomy 149
4.3.3 The Performance Expectations Taxonomy, Bloom's Taxonomy of Educational Objectives and modifications of Bloom's Taxonomy of Educational Objectives 153
4.3.3.1 *Conceptual relationships between the taxonomies* 154
4.3.3.2 *Calibration of the taxonomies* 157
4.4 **The relationship between the objectives and aims of the Senior Certificate Core Biology Syllabus and the Performance Expectations Taxonomy** 164
4.5 **Examples of Senior Certificate Biology examination questions classified using the Performance Expectations Taxonomy** 166
4.6 **Chapter summary** 167

CHAPTER 5 **METHODS FOR THE ANALYSIS OF QUESTION PAPERS AND CANDIDATES' ANSWER SCRIPTS** 179
5.1 **Data sources** 182
5.1.1 Senior Certificate Biology examination question papers 182
5.1.2 Candidates' answer scripts 183
5.1.3 The representivity and reliability of the sample of candidates' answer scripts 190
5.2 **Source analysis** 193
5.2.1 Examination question papers 194
5.2.1.1 *Complete examination question papers*.......... 195
5.2.1.2 *Item-by-item categorization* 196
5.2.1.3 *Profile of complete examination question papers* 200
5.2.2 Candidates answer scripts 203
5.2.3 Descriptive and analytical methods 204
5.2.3.1 *Univariate analyses* 204
5.2.3.2 *Comparison of content* 204

5.2.3.3 *Multivariate analyses* 208
a **_Cluster analysis_** 208
b **_Internal consistency of examinations_** 209
c **_Discriminant function analysis_** 210
d **_Factor analysis_** 210
5.2.3.4 *Difficulty of examinations* 211
5.3 **Chapter summary** 213

CHAPTER 6 **ANALYSES OF QUESTION PAPERS AND CANDIDATES' ANSWER SCRIPTS** 215
6.1 **Analysis of question papers** 219
6.1.1 Structure of question papers, structure of questions and structure of answers 220
6.1.1.1 *Macrostructure of question papers* 221
a **_Number, types and length of question papers_** 222
b **_Total marks of examination_** 228
c **_Sub-sections of question papers_** 228
6.1.1.2 *Structure of questions and free response answers* 231
a **_Structure of questions_** 232
b **_Structure of free-response answers_** 246
6.1.2 Content standards 256
6.1.2.1 *Topics* 257
6.1.2.2 *Performance expectations* 262
6.1.2.3 *Content (topics, performance expectations)* 269
a **_Similarity of content_** 286
b **_Breadth and depth of knowledge_** 296
6.1.3 Comparison of profiles 300
6.2 **Analysis of answer scripts** 306
6.2.1 Performance according to structure of question paper 307
6.2.1.1 *Choosing the correct answers or free-response answers* 308
6.2.1.2 *Length of questions* 308
6.2.1.3 *Use of non-text elements in questions* 313
6.2.1.4 *Length of answers* 316
6.2.1.5 *Non-text elements required in answers* 316
6.2.2 Performance standards 320
6.2.2.1 *Performance according to topic* 320
6.2.2.2 *Performance according to performance expectations* 323
6.2.2.3 *Performance according to content* 333

6.2.3 Reliability of student performance data ... 345
6.2.3.1 *Internal reliability* ... 345
6.2.3.2 *Factor analysis* ... 354
6.2.3.3 *Classification of students by aggregate* ... 358
6.3 **Difficulty of examinations** ... 370
6.3.1 Question papers ... 371
6.3.2 Student performance ... 373
6.4 **Equating of examinations** ... 384
6.5 **Optimal length of examinations** ... 389
6.6 **Chapter summary** ... 395

CHAPTER 7 **CONCLUDING CHAPTER** ... 399
7.1 **Providing answers to the research sub-questions** ... 401
7.1.1 Research sub-question 2: What were the profiles of SC Biology examination question papers in terms of what they assessed directly? (1994 to 2007) ... 402
7.1.1.1 *Structural characteristics of question papers* ... 402
7.1.1.2 *Content standards of questions papers* ... 404
7.1.2 Research sub-question 3: What were the differences between HG and SG SC Biology examination question papers? (2001 to 2007) ... 406
7.1.2.1 *Structural characteristics of question papers* ... 406
7.1.2.2 *Content standards of question papers* ... 408
7.1.3 Research sub-question 4: How did the profiles of SC Biology examination question papers change during the specified period, if at all? (1994 to 2007) 409
7.1.3.1 *Structural aspects of question papers* ... 410
a ***1994 and 1995 HG*** ... 410
b ***1996 to 2000 HG*** ... 410
c ***2001-2007 HG and SG*** ... 411
7.1.3.2 *Content standards of question papers* ... 412
a ***1994 and 1995 HG*** ... 412
b ***1996 to 2000 HG*** ... 413
c ***2001-2007 HG and SG*** ... 414
7.1.4 Research sub-question 5: How did candidates of various total mark categories within the various SC Biology examinations compare in terms of the kinds of questions they could and could not answer successfully? (2005 and 2006) ... 415
7.1.4.1 *Structural aspects of question papers* ... 415
7.1.4.2 *Content standards of question papers* ... 417

7.1.5 Research sub-question 6: What relationships, if any, characterized the achievements of the same candidates writing both Paper 1 and Paper 2 of the SC Biology examinations? (2006) 418

7.1.6 Research sub-question 7: What relationships, if any, emerged between the profiles of SC Biology examination question papers and student achievement? (2005 and 2006) 419

7.2 **Providing answers to the research question** 420

7.3 **Limitations of this study** 423

7.4 **Significance of this study** 425

7.5 **Recommendations for practice and future research** 426

7.6 **Concluding remarks** 429

REFERENCES 435

VOLUME II

APPENDICES 485

CHAPTER 1

INTRODUCTION

In general, examinations signal two messages to the society in which they operate: first, the competencies that are valued, that is, its standards; and second, the required level of mastery in these competencies that are indicators of success. In South Africa, Senior Certificate (SC) examinations, offered in a number of subjects at the end of Grade 12, signify the end of secondary schooling. The SC serves as both a school-leaving certificate and a university entrance qualification (National Education Policy Investigation [NEPI], 1992a; Trümpelmann, 1991) thus shaping the future opportunities of individuals. Therefore, the quality of the data from assessments like the SC that is used to make decisions about students at the end of their high school careers should be of utmost importance (Cone & Foster, 1991). Investigating and improving the usefulness of such data, in line with current measurement practices, should thus be the joint concern of measurement specialists, test users and the consumers of test information (Cizek, Bowen & Church, 2010).

The SC was not part of an explicit standards-based curriculum, so there is little understanding about the relationship between student achievement in the SC examinations, competency and standards. Both in South Africa and internationally, 'standards' is an often-used educational term, the meaning of which has been confused in the literature and by public use. In this study, a methodology to make explicit the standards inherent within the SC Biology examinations—and the relationship between standards and student achievement—was developed, described and applied. The methodology involved using the international research literature about educational standards for the construction of a conceptual framework by which the implicit standards of the South African SC Biology examinations could be extracted and understood.

The conceptual framework developed in this study uses a reciprocal relationship between standards and validity evidence where different types of validity evidence are used to explicate the standards inherent in the SC Biology examinations. This framework draws critically on the understanding, and use of educational standards globally, especially in the United States of America (USA) and in the United Kingdom (UK). Effective application of the conceptual framework for use in explicating the standards of SC Biology examinations required the development of an instrument to determine the cognitive demand of questions.

Retrospective document analysis was used to study four different kinds of documents: Biology curriculum policy documents; South African SC Biology examinations over a period of 14 years; students' SC Biology examination answer scripts over two of the 14 years; and government records to generate the context of this study and the qualitative and quantitative descriptions of standards elicited from these documents. The study's findings are related to past and current assessment, especially examination practices internationally and in South Africa.

This study does not constitute a judgment about the quality of the SC Biology examinations. Instead, it was intended to illustrate the use of the methodology developed to explicate standards of the SC Biology examinations from 1994 to 2007 and to use the standards thus generated comparatively. This study is an attempt to start a broader conversation, and critical reflection, about how we gather validity evidence, an indicator of standards, to support the inferences that we make about Biology learners' knowledge in South Africa. Inferences made from SC examination results are high-stakes because of how they affect the future lives of those who take the examinations. It is therefore important that equity and fairness be ensured for all students who write, in this case the Biology, SC examinations. While drawing on international theoretical and practical assessment approaches, the conceptual framework and the methodological aspects of this research are framed by the South African practice and what the author believes is possible given the current infrastructure and funding allocated to assessment in the country.[1] It is hoped that the assumptions and findings of this study enable reflection, and will in the future be debated, tested and challenged by, possibly the use of more sophisticated psychometric analyses in order to refine our SC examination practices in all subjects. Particularly we must better understand the student performance in the SC with respect to standards.

1.1 Background and rationale

> *"A nation's educational standards are embodied in its secondary-school-leaving examinations"* (Eckstein & Noah, 1993, p. 143)
>
> *"[A]cost of not having national standards is the cost of non-comparability"* (Noah, 1989, p. 18).

[1] Throughout this thesis, the author has tried to remain mindful of critiques of international comparisons of educational standards which have argued rather that studies about educational standards should be conducted *in situ* so that they provide relevant information that can be used by educators and policy makers in their own countries (Resnick, Nolan, & Resnick, [1996] cited by Louis & Versloot [1996] and Schwille [1996]).

Post-1994, with the dawn of democracy in South Africa the new government made a commitment to ensure quality education for all South Africa's citizens (African National Congress [ANC], 1994; Department of Education [DoE], 1995). For South Africa, the adoption of "education for all" has meant a major restructuring of the unequal education system that was inherited from the pre-1994 (apartheid) government. A number of changes to the education system are relevant to this study. For example, new examining authorities replaced old ones; pre-existing examining bodies were merged; interim syllabuses were introduced (e.g., DoE, 2001a; Western Cape Education Department [WCED], 1995a, 1995b, 1996); the formats of examinations changed; and a formative assessment mark as part of examination scores was introduced. Almost 20 years after the ANC's promise that all South Africans will receive a quality education there is a continuing concern amongst all sectors of South African society, that the quality of education provided by the government, especially that of high-school leaving examinations, has continued to decline (Jansen, 2012).

One collection of externally set and marked examinations at the end of Grade 12─the SC─was meant to certify that candidates were competent to take their place as responsible citizens and/or to enter the workplace, or were eligible to qualify for admission to tertiary education (a matriculation endorsement) (Ndhlovu, Sishi & Deliwe, 2006; NEPI, 1992a). These examinations signified the acquisition of predefined content and 'mastery' of particular subjects. While the content which students were required to know was stated in the syllabus and associated guideline documents, we do not know exactly what 'mastery' of a subject meant (i.e., what competencies these examinations measured or did not measure). Without this information, we cannot interpret students' achievement in the examinations (Nitko & Brookhart, 2007), nor can we begin to understand what changes in student achievement between years might mean.

The dual purpose of the SC examinations, that is, to serve as both a terminal certification of successful completion of high school as well as a prognostic test for university admission, has been a source of debate from as early as the 1920s in South Africa (Trümpelmann, 1991) and continues to be contested (Lolwana, 2006). The South African national SC pass rate, that is, the proportion of students who qualified for a SC, decreased from 54% in 1996 to 47% in 1997. Thereafter, it steadily increased to 73% in 2003 (Naidoo, 2006; wa Kivilu, 2006), and in 2004 it started to decrease again (Fiske & Ladd, 2006) reaching 65% in 2007 (DoE, 2009a). During a similar period of time, Yeld (2005a) noted steadily lower graduation rates,[2] and high drop-out rates in many South African universities. More recent figures give South Africa's university graduation rate of 15% as amongst the lowest in the world (Letseka & Maile, 2008).

There was much speculation about the reasons for the variable SC pass rate. It was unclear whether the increase in pass rate was a result of a corresponding improvement in the quality of schooling or a lowering of standards (Zille, 2005), or the fact that a higher proportion of students passed on a lower level than on a higher level (Lolwana, 2006). There was no formal policy that required the use of these examinations for accountability or monitoring purposes, yet the South African government used the increasing pass rates as evidence of an improving education system (e.g., Government Communication and Information System, 2005/2006; News24.com, 2003). Moreover, the public continued to want to use the SC results to assess how well their school-leavers are doing (Chisholm, 2004a) and the quality of their education (Taylor, 2009). Such use of the SC results continued despite a mistrust of the value of SC examination question papers and indeed of the examination process as a whole (Naidu, 2006), and despite little evidence on which to evaluate whether increased pass rates implied improvement, or no improvement, in what was learned by students at the end of the SC year. Universities attributed the recent increases in first-year failure rate to a lowering in the standards of the SC examinations and, in part, to a decline in the level of cognitive challenge of these examinations (Yeld, 2005b), which is sufficient to be "a threat to the learning health of the [South African] nation" (Muller, 2005, p. 44) .

"Assessments communicate values, standards and expectations" (Mislevy, Steinberg & Almond, 2003, p. 4); they influence peoples' understanding of what is important to learn and what learning is (Moss, Girard, & Haniford, 2006); and have been perceived to be crucial to the improvement of teaching and learning (Linn, 1993). Therefore, assessments, like examinations, are potentially "powerful instruments for educational change" (Shavelson & Huang, 2003, p. 11), provided their results can be, and are interpreted in a meaningful way. In South Africa, speculative interpretations with little empirical evidence as to the meaning of student achievement parallel what happened in the USA where "test results have become the dominant way states, politicians, and newspapers describe the performance of school" (Dorn, 1998, p. 2). Similarly, in England (Spencer, 2003, p. 135) where "performance tables on which parents and others judge schools" are used, Black (2002, Section 3.2, para. 3) cautions that, in the absence of an understanding of assessments, "the

2 The graduation rate refers to the number of students who start and successfully complete university study.

unsophisticated make inferences, are then disappointed, and critici[z]e the qualifications". In South Africa, opinions as to what improved pass rates mean are based on little or no data as to the quality of examinations or how they relate to student achievement. Such a lack of understanding of the relationship between examinations and student achievement has led to those charged with evaluating and deciding assessment policies possibly undervaluing (or overvaluing) the important information provided by what the SC examinations can tell us about student learning. The lack of understanding also makes determining the quality of these examinations and establishing their comparability between years difficult. In the Netherlands "[k]eeping up standards [in national secondary school examinations] is therefore essential for maintaining the confidence of tertiary education in the performance level of incoming students" (Alberts, 2001, p. 353) and mechanisms to equate examinations between years are necessary to achieve this maintained confidence.

South Africa does not offer a standards-based curriculum, and so what competencies the SC examinations measure or do not measure, in terms of standards, is unclear. Without information about competencies, we cannot interpret students' achievement in the examinations (Nitko & Brookhart, 2007), nor can we begin to understand what changes in student achievement between years might mean. "There exists no common public understanding of what standards our [South African] learners are expected to achieve" (wa Kivilu, 2006, p. 34). As the "cost of not having national standards is the cost of non-comparability" (Noah, 1989, p. 18) it is impossible to consistently gauge standards over time.

"Secondary school-leaving examinations articulate a nations aspirations … [and t]hese high-stakes examinations are important articulations of national policy with significant social and economic consequences" (Valverde, 2005, p. 52). Umalusi (2004), as the quality assurer of standards in the South African education system, and the *Marking Matric* project (Reddy, 2006a), have indicated the need for a better understanding of SC examinations results and, more specifically, what they can reveal about the state of education in South Africa. If "a nation's [South Africa's] educational standards are embodied in its secondary-school-leaving examinations" (Eckstein & Noah, 1993, p. 143), the examinations functionally define what we value in learning, and we need to begin to understand what information about standards our SC examinations convey to society. Only by articulating and analyzing the standards of the SC examinations can we begin to understand what our expectations of schools are, whether these expectations are reasonable, what expectations need re-thinking, and what the implications of an understanding of standards are for the formulation and development of education policy.

In recent years, much reform effort in the area of schooling in South Africa has focused on the introduction of a new outcomes-based education system (DoE, 1997a) embodied in a newly developed National Curriculum Statement (NCS) (DoE, 2000a), and the first National Senior Certificate (NSC) Grade 12 examinations were held in 2008.[3] Two authors have argued that the results of the SC and the NSC cannot be contrasted because the "two systems [SC and NSC] cannot be directly compared as they work from two different sets of assumptions" (Govender, 2009, p.2) and the NSC was "new … preventing any comparison with previous years" (Hindle, 2009, p. 8). Choppin (1981, p. 1) argued that because educational systems change with time, it is crucial that changes be explicated in a way that allows comparisons to be made, if society is to understand whether "these [educational] changes are in the right direction". If we are to attempt to begin to understand the impact of the NCS on teaching and learning, especially the role of the 'new' NSC in this curriculum, we need to have a point of reference. That is, we need a benchmark and an empirically supported understanding of what the 'old' SC practice was, against which the new can be compared (Chisholm, 2004b; Dickson, 2005). Taylor (2009, p. 21) has called on critics of the new curriculum and its assessment practices to bring "a more informed and rigorous analysis of the available data" to the debate about the meaning of the SC and the NSC examination results.

The issues outlined above indicate the necessity for an understanding of South African SC examinations, based on empirical evidence, especially with regard to standards. The absence of well-articulated, and commonly understood, educational standards in the South African context of SC examinations necessitates an explicit unfolding of how standards in this context can be understood. While it is important to acknowledge that three factors, that is, individual effects, school effects, and the examination paper and syllabus effects, influence student achievement in examinations (Jones & Ratcliffe, 1996), this study focuses on an analysis of examinations and their accompanying policies, in the context of South African practice and international educational standards practice.

[3] SC Biology was replaced by NSC Life Sciences

1.2 Problem statement

In South Africa, SC examinations comprise question papers, student answer scripts and student achievement marks but there is a lack of understanding as to how these components relate to each other, within and across years. Hence there is no explicit description of what different levels of student achievement imply in terms of student competencies or educational standards. Designing appropriate examinations should be based on valid and reliable data about both how examinations measure─and can measure─student competencies, and what student achievement in examinations might mean. Badly designed examinations undermine student learning (Amrein & Beliner, 2002). Trustworthy validity evidence about examinations is necessary to assure that the scores obtained by students taking assessments are useful, meaningful and defensible (Downing & Haladyna, 1997; O'Niel, Sireci & Huff, 2003-2004; Sireci, 1998; Sireci & Parker, 2006). Establishing the validity of an assessment, by using validity evidence, should occur each time a new or modified assessment is used (Messick, 1989a, 1989b). In South Africa there are no existing processes or frameworks by which the validity of the SC examinations or educational standards can be argued or established. Hence, to date, validity is either ignored or implicitly assumed as a vague subjective criterion to which the SC examinations generically comply.

1.3 Aim

The aim of this study is to develop, to describe, and to apply a methodology to generate the conceptual and empirical evidence necessary to explicate and compare the standards of SC Biology examinations. The purpose is to understand what can be inferred from, or explained by, this evidence about retrospective plausible student competencies as students left the secondary school system.

1.4 Objectives

To achieve the specified aim the study has the following objectives:

1. To construct a relevant and meaningful explanation of educational standards and their role in curriculum, especially examinations.

2. To develop a conceptual framework by which educational standards can be extracted and made explicit from SC Biology examination question papers and from marked candidates' answer scripts respectively, where available. In this conceptual framework validity evidence is used as a proxy for possible but implicit standards, and then standards are used to describe what facts, concepts and skills examinations are assessing and how students are performing.

3. To develop a methodology by which educational standards, as conceptualized in objective 2 above, can be elicited *post hoc* from South African SC Biology examination question papers and candidates' answer scripts.

4. To ensure that South African examination practices remain the focus of this work while drawing on international assessment and measurement practices, by using a 'language' in the conceptualization and operationalization of the framework that is rooted in South African policy and practice.

5. To apply the methodology rigorously to SC Biology examination question papers covering a period of fourteen years, 1994 to 2007, and a selection of marked candidates' answer scripts for two of these years, 2005 and 2006.

6. To describe, analyze and interpret patterns and changes in the educational standards extracted from Biology SC examination question papers and candidates' answer scripts of the selected period.

7. To discuss the prospective implications of the findings of 1 to 6 on NSC Life Sciences examinations policies and practices in South Africa.

1.5 Research question and sub-questions

1.5.1 Research question

The main research question of this study is:

What did the SC Biology examinations in South Africa assess; did their focus change during the period 1994 to 2007; and, if so, what did this change mean?

This research question is broken down into seven research sub-questions, each of which addresses a particular aspect of these examinations. Sub-question 1 concerns the development of a conceptual framework which enables both the examination question papers and the available candidates' answer scripts to be viewed through the lens of educational standards. Sub-questions 2 to 4 address issues related to the only the examination question papers, while sub-questions 5 to 7 pursue issues concerning both the question papers and the candidates' answer scripts. Sub-questions 3, 5, 6 and 7 in particular explore claims made about the examinations, which were used as a basis to certify candidates at different levels of competency, according to (HG) or Standard Grade (SG).

1.5.2 Research sub-questions

Standard Grade question papers and HG and SG candidate answer scripts were not available for the entire period 1994 to 2007, despite diligent efforts to obtain complete sets of papers and suitable exemplars of associated candidates' answer scripts. Consequently, where appropriate, a specific timeframe for each sub-question is given in brackets, after each question.

1. What are *educational standards,* and how might they be used to describe and compare SC Biology examination question papers and candidates' answer scripts? (1994 to 2007)

2. What were the profiles of SC Biology examination question papers in terms of what they assessed directly? (1994 to 2007)

3. What were the differences between HG and SG SC Biology examination question papers? (2001 to 2007)

4. How did the profiles of SC Biology examination question papers change during the specified period, if at all? (1994 to 2007)

5. How did candidates of various total mark categories within the various separate SC Biology examinations compare, in terms of the kinds of questions they could and could not answer successfully? (2005 and 2006)

6. What relationships, if any, characterized the achievements of the same candidates writing Paper 1 and Paper 2 of the SC Biology examinations? (2006)

7. What relationships, if any, emerged between the profiles of SC Biology examination question papers and student achievement? (2005 and 2006)

1.6 Clarification of terms

Terms for which the meaning is constructed, or argued in this thesis, are not listed below. The meanings of some of these terms are constructed in this thesis, in which case the relevant chapter is given in brackets after the explanation .

Assessment	The process for obtaining information that is used for making decisions about students, curricula, programs and education policy (Nitko & Brookhart, 2007).
Candidate	A person who takes an examination, sometimes called a student.
Examination	Tests that are administered and scored according to highly structured, prescribed directions. Each examination consists of a question paper (s) and candidates' answer scripts.
Content standards	Define what students should know, and be able to do with what they know (Chapter 3)
Examining Body	An institution that sets and/or administers and /or marks examinations.
Government	Administrative bureaucracy which controls the state, South Africa.
Grade	(1) Capital 'G' refers to the Grade on which SC Biology was studied, i.e., Higher Grade or Standard Grade ; (2) the year of school study, e.g., Grade 12 refers to the final SC year (from 1996 onwards); (3) lower case 'g' also refers to marks or symbols.
Higher Grade	One of two levels on which a student could study a SC subject – the other level is Standard Grade.
Mark	Examination questions assigned units called marks.
Marking	Awarding marks to candidates answers.
Matriculation Endorsement	Students who achieved their Senior Certificate with particular combinations of subjects, and who achieved requisite marks in these subjects were awarded a Senior Certificate with Matriculation Endorsement which made them eligible for tertiary study at South African universities.
Measurement	Procedure for assigning numbers, or scores, to a specified attribute or characteristic of a person so that the number describes the degree to which a person possesses the attribute (Nitko & Brookhart, 2007).

Memorandum	The South African name for marking guidelines.
National Senior Certificate	National Senior Certificate replaced the Senior Certificate from 2008 onwards.
Performance expectations	What students are expected to do with their subject matter knowledge (Chapter 3).
Performance standards	Performance levels which span a performance continuum of different performance levels, each of which is defined by differential performance in the content standards (Chapter 3).
Profile	A combination of the content standards and structural characteristics of an examination question paper (Chapter 5).
Quality	Is denoted by standards (Chapter 3).
Reliability	The extent to which measures of tests are consistent (Nitko & Brookhart, 2007) (Chapter 3).
Scorable event	The smallest [discrete] questions in an examination that cannot be broken down into more sub-questions (Chapter 5).
Score	Total number of marks awarded for a defined task.
Senior Certificate	Certificate awarded to successful students at the end of their last year of high or secondary school. Students who fulfil certain requirements receive a Matriculation Endorsement.
Standards	Content standards and performance standards (Chapter 3).
Standard X	In South Africa, “Standard” was the name given to a particular year of school study prior to 1996. Years 1 and 2 were respectively called Sub A and Sub B. Year 3 was called Standard 1, and Year 12 was called Standard 10, and was referred to the Senior Certificate year.
Standards-based curriculum	Teaching and assessment are determined by the same clearly defined, articulated and explicit standards.
Standard Grade	One of two levels on which a student could study a SC subject – the other level is Higher Grade.
Symbol	Symbols A to H represent descending levels of student performance also called the grade.
Task	Any instruction, given to students to elicit a response or answer from the student.
Test	An instrument or systematic procedure used to collect evidence to make inferences about a student’s learning .

Validity	"[T]he degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (Messick, 1989, p.13) (Chapter 3).
Validity evidence	Evidence that is systematically collected about an assessment to integrate the inferences made about students from their performances therein (Chapter 3).

1.7 Organization of the remainder of this thesis

This thesis does not follow the traditional structure that many theses have. The approach to this thesis was essentially eclectic, because it draws on a variety of different perspectives about the nature of educational standards and of assessment in general. The 'story' which this thesis tells in addressing the research question posited in this chapter develops sequentially from Chapters 2 to 7. Therefore, there is no specific literature review – the literature review is interwoven into all of the chapters. The delimitations of this study are not declared in Chapter 1 – instead they are given at the end of Chapter 2, which is a description, compiled from a number of disparate sources, of the unique context in which the SC examinations in South Africa were conducted. There is no specific methods section because this study uses a combination of methodological approaches which are constructed and appear in different chapters. A variety of methodologies is used to show what is possible within a South African context. Aspects of methodology appear in Chapter 3 where the conceptual framework is developed to guide, in terms of standards, the analysis of question papers and candidates' answer scripts which are the focus of the research question. A further aspect of methodology appears in Chapter 4 where current methods of measuring one of the features of standards identified in Chapter 3, cognitive demand, are reviewed and questioned. An instrument is developed and validated specifically for use as a measure of cognitive demand in this study. Chapter 5 describes the methodology by which the standards conceptualized in the framework given by Chapter 3 are operationalized. Chapter 6 describes the standards of the SC examinations analyzed in this study. Elements of results, discussion and critique appear in each of Chapters 2 to 6.

Each of the Chapters 2 to 7 begins with a figure that links antecedent chapter contents, by bold arrows, to the shaded area in which the concerns of each new chapter are summarized. Each of the Chapters 2 to 6 concludes with a short summary of that chapter.

Another way in which this thesis differs from many other theses is that the research question and all but one of the sub-research questions (research sub-question 1) are only directly addressed in Chapter 7, the last chapter of the thesis. Research sub-question 1 is addressed in Chapters 3, 4 and 5 (Objectives 1, 2, 3 and 4). Chapters 2, 3, 4 and 5 need to be presented to the reader before the empirical evidence, generated in Chapter 6 (Objectives 5 and 6), can be used to directly answer the remaining research sub-questions 2 to 7 and the research question in Chapter 7. Recommendations for the examination of NSC Life Sciences (Objective 7) are made on the basis of the empirical data presented in Chapters 2 to 6. To help readers navigate the relationships between the objectives, the research question and research sub-questions and the various chapters are shown in Figure 1.1.

The nature of the notion of standards is conceptualized (Chapter 3) and operationalized (Chapters 4 and 5) to elicit, describe and compare the Biology SC examinations over a period of fourteen years necessitated a number of lengthy tables and figures. Their purpose is to capture the sometimes subtle changes which were observed in standards within and between years. Wherever appropriate the author flags connections, or uses summaries, to navigate the reader between the many tables, figures and appendices.

Post-1994 South African education policy documents which introduced the new outcomes-based education (OBE) curriculum[4] and which emphasized the use of new terminology, have been described as containing a "jungle of new jargon" (Vandeyar & Killen, 2003, p. 125). The new terminology was not always understood or used by everyone in the same way and this divergence led to continued confusion and discontent, especially amongst teachers (Dada, Dipholo, Hoadley, Khembo, Muller & Volmink, 2009). To avoid confusing the reader of this thesis with respect to the nuances of the South African educational context as well as the distinct ways in which terminology is used in the international literature, the author deliberately makes extensive use of footnotes. The footnotes serve either to orientate the reader and/or to provide explanatory notes which might interfere with the flow of the text if inserted into the text itself.

A number of specific examples of questions from the SC examinations analyzed in this study are presented within this study (Chapters 4 and 6) as it became clear to the author during the research that many of these question papers have not been archived, not even by extant examining bodies, and are therefore not available for all readers to consult.

[4] The OBE curriculum, introduced in 2006 at the Grade 10 level, was examined for the first time in the 2008 NSC examinations.

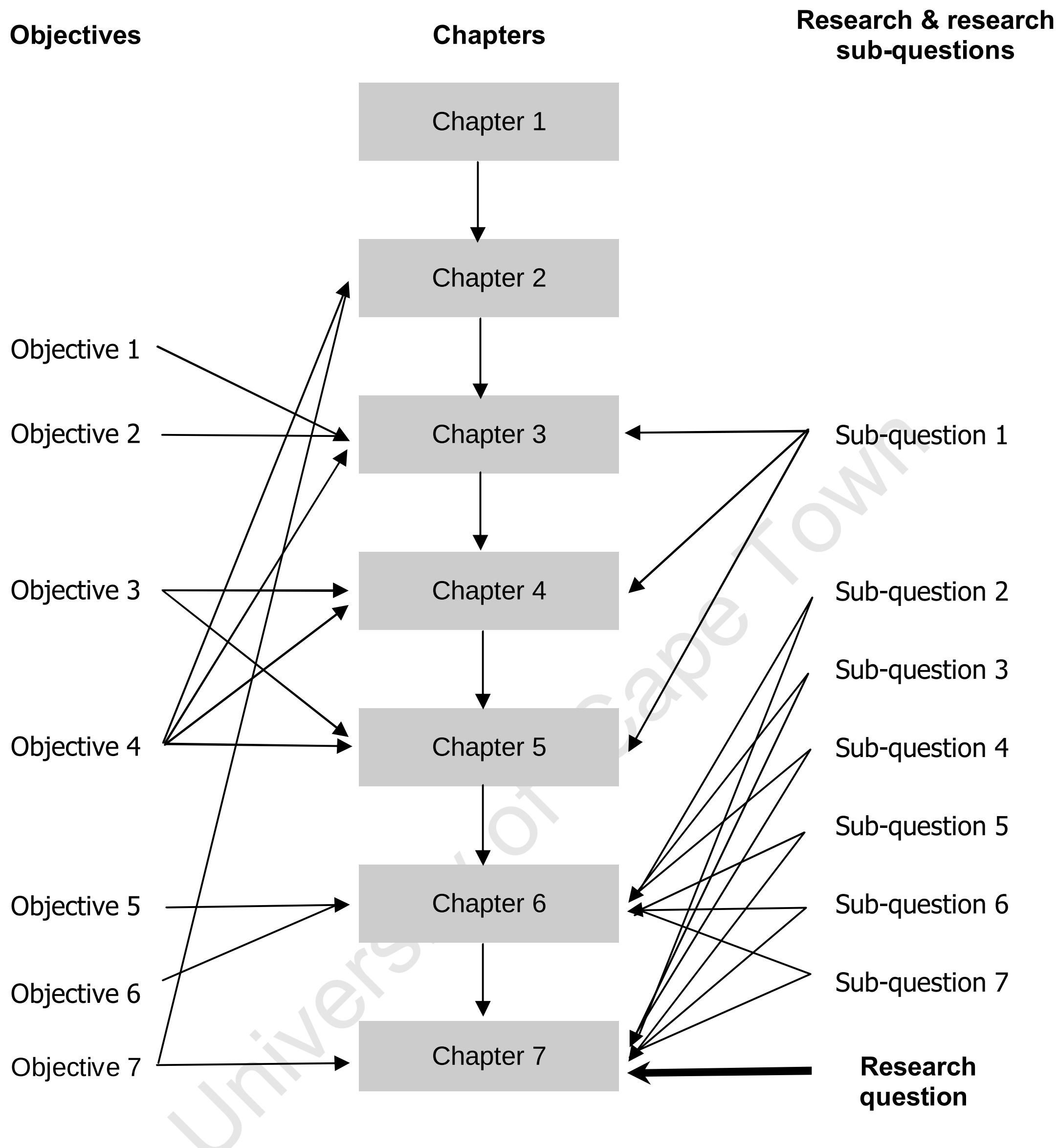


Figure 1.1 Structure of thesis showing the relationships between the chapters, objectives, sub-research questions and the research question.

After reading Chapter 1, the reader might like to preview the limitations of the study and the significance of the study which are addressed in Chapter 7. Given the potential scope of this thesis, and the vast international literature about assessment, the author seeks to acknowledge explicitly the limitations of her approach in this research so as to not create false expectations for the reader.

A synopsis of the 'story' as told by this thesis follows here. In Chapter 2 a review of the contextual factors related to this study is given. This chapter includes descriptions and discussions of the policies which guided the SC Biology examination practices during the period of this study. Normally in a thesis the areas of the research focus, that is, the delimitations or scope of the research would be explained in Chapter 1, together with the problem statement and the research question and research sub-questions pertaining to the study. During her research the author realized how different the South African context, especially the high school leaving examination process, was from corresponding practices elsewhere in the world. Therefore, readers of this thesis would benefit from first being offered an understanding of the nuances of the local context, in order to appreciate the areas of focus selected for this research.[5] For this reason the areas of research focus follow on from descriptions of the context in Chapter 2. The scope of this research was of necessity practically dictated by available sources of question papers and answer scripts rather than by design. Chapter 2 includes a short summary of some international comparative studies of science assessment systems, some of which make at least some implicit mention of standards.

Chapter 3 begins by developing a conceptual understanding of the global use of standards in the context of education. This element is followed by a review of how standards are used by the research community in different countries to understand their curricula. The development of the conceptual framework to extract explicit standards (both content standards and performance standards) from South African SC examinations is then described. This framework uses standards as a lens for the study and validity evidence as a proxy for standards. Four aspects of validity evidence are identified in this framework as being necessary to generate the content standards. Three of these aspects, the general structure of a question paper, the structure of the questions and of the answers required and the topics covered are considered as objective since they can be reliably categorized according to checklists. The fourth aspect, cognitive demand, can elicit very subjective categories because of the variety of taxonomies that researchers use to classify the performance expectations of assessment items.

5 The archiving of material pertaining to the South African SC Biology examinations was not always complete and what has been archived is not necessarily readily accessible to researchers. Therefore, the author has pieced together as much material as she could find to build an objective record of activities associated with the South African SC biology examinations, against which to position this study.

In Chapter 4, the development and validation of the instrument or taxonomy used to measure performance expectations in this study, is described. The developed instrument is calibrated against the taxonomy that was used by the DoE, when it set some of the Biology question papers.

In Chapter 5, the methods of data collection and the methods used to explicate and describe the standards implicit in the question papers and candidates' answer scripts, that is, the operationalization of the conceptual framework are discussed.

In Chapter 6, the results of the analyses of data from the question papers and candidates' answer scripts are presented. The results are organized separately for the examination question papers (the content standards) and for the candidates' answer scripts (the performance standards). The volume of results, especially the descriptions of the standards generated from the question papers and scripts necessitated condensation of the results into many figures, tables and appendices. These summaries are connected at the beginning of the chapter to the accompanying text.

In Chapter 7, the content standards and the performance standards generated in Chapter 6 are used to address each of the research sub-questions 2 to 7, and the research question. What follows thereafter is a discussion of the limitations of this research. An argument is made for the significance of this study in contributing to an international understanding about assessment, particularly about standards, and to an understanding of the policies and practices which operated in South African SC Biology examinations, from an international perspective. Recommendations are offered for further research in the area of SC/NSC assessment in South Africa. A series of questions about SC/NSC assessment practices in South Africa generated by this study may constitute the basis of further research. The author's concluding remarks complete this chapter, and the thesis.

To facilitate the reading of the numerous appendices concurrently with the relevant text, this thesis has been bound as two volumes. Volume I contains the Chapters 1 to 7 together with the References, and Volume II contains the Appendices.

CHAPTER 2

CONTEXTUAL FACTORS

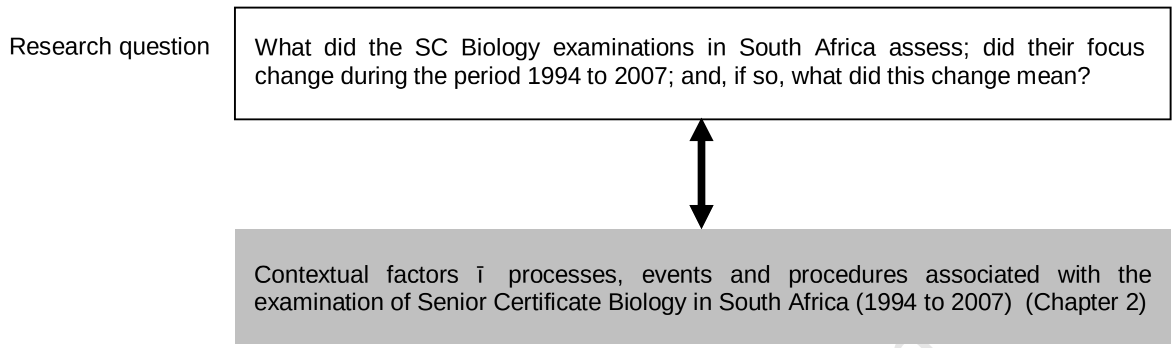


Figure 2.1 The relationship between Chapter 2 and the research question stated in Chapter 1.

An assessment represents an interrelationship between its intended purpose and the context in which it is used (Mislevy, 1995). The design and implementation of assessments takes place within contexts that are political, social, institutional, historical (McNamara, 2003), as well as economic and ideological (Keeves, 1994). Therefore, in one-country studies aimed at improving an assessment system it is insufficient to look only at general trends − understanding the particular, the specific and the cultural of the national context is crucial (Kellaghan, 1992; Little, 1996). Given the unequal spread of technical expertise in crafting assessments across the world, transfers of assessment practices take place between countries (Sebatane, 2000). Noah (1996, p. 94) highlighted that the level of 'copying' or 'borrowing' or similarity of assessment practices between countries was highly variable because "national idiosyncrasy, is and has been, the rule". More recently, Luke (2011, p. 374) advocated "principled [educational] policy borrowing" between countries, which requires an intimate and careful understanding of the contexts of both the borrowing and the lending countries. Therefore, in order to understand how the research question posed in Chapter 1 might initially be answered by critically using the international literature on assessment, it is necessary to describe the 'idiosyncrasies' or contextual features of the South African educational system and those of the countries from which the literature is drawn.

In apartheid South Africa, the white-minority government deliberately legislated to ensure disparity in the quality of education offered to people of different populations groups,[6] over a long period of

[6] Population groups were defined by the government of the time. These groups, as they apply to the SC examinations are explained in Section 2.1.2.

40 years. This policy created “racial and class inequalities in the distribution of education resources and facilities” (NEPI, 1992a, p.9) which remain in the education system today. In 1994 the installation of a new democratically elected government heralded the start of a period of transition and transformation. Ndlovu, Sishi and Deliwe (2006) identified three phases of this transition with respect to education assessment reform in South Africa, namely, 1994 to 1998, 1999 to 2003, and a phase starting in 2004. For SC purposes, the third phase could be considered to have ended in 2007, the last year in which the national SC examinations using the Nated 550 syllabuses[7] were written. This study recognises three different periods, 1994 and 1995, 1996 to 2000 and 2001 to 2007, in the examining of Biology at SC level, based on the differing examining bodies[8] which operated during these periods and the policy requirements which prevailed during these periods. These changes in examining bodies and examinations policy requirements are discussed in this chapter.

“Difference[s] between [South African] students’ performance in a national exam has much more to do with the school they went to and their socioeconomic status and the conditions under which they live and learned” (Price, 2010, p. 13) than just comparing the marks each achieved. Examples of inequities that have persisted into the current South African education system are summarized in Table 2.1. While this study does not attempt to address the social and educational disparities that existed in South African society during the period of study, the readers should bear these inequities in mind as they read about the policies and practices which prevailed at different times of this study.

The first part of this chapter is a description of the processes, events and procedures associated with the SC and its examinations in South Africa. The second part of this chapter outlines the reasons for the selection of Biology (subject), Grade 12 (level of schooling), the time-period 1994 to 2007 and the use of candidates’ answer scripts from only the WCED, as the foci of the student performance research for this study. The author has made a deliberate attempt to provide a rich and detailed description of the context of this study, to provide readers with an accurate and intelligible sense of the complex context in which the SC Biology examinations operated.

---

7 The pre-2008 Grade 12 curricula and their accompanying syllabuses have become commonly collectively referred to as the Nated 550 curricula and syllabuses. Considering these documents collectively is not correct as there were four different versions of Nated 550 which operated during the period 1994 to 2007. Each version of the document had different policy implications for the teaching and examining of Biology at SC level. These versions of Nated 550 are discussed in Section 2.2 of this chapter.

8 Also known as examination bodies, examination authorities and examining authorities in South Africa.

Table 2.1 A summary of different education statistics by province (government schools only), 2008/2009. Compiled from The Children's Institute (2010).

Province	Number of public schools	Proportion of schools (%)				Number of school-age children (7-17 years)	Proportion of school-age children (%)			
		No-fee status	More than 45 learners/ classroom	Sanitation	Library space		Attending school	Access to a school	Completed Grade 9	Achieved 50% benchmark for literacy (Grade 6)
Western Cape	1 451	45	11	98	25	962 000	96	95	73	63
Northern Cape	599	66	9	86	12	264 000	96	82	61	56
Gauteng	1 990	51	23	95	18	1 620 000	96	87	74	51
Free State	1 681	86	8	59	9	695 000	98	84	62	27
KwaZulu-Natal	5 877	58	30	57	6	2 556 000	96	71	63	25
North West	1 730	82	16	62	6	746 000	96	67	60	22
Mpumalanga	1 893	57	31	69	6	916 000	98	78	55	22
Eastern Cape	5 723	89	33	39	3	1 859 000	96	74	48	15
Limpopo[a]	4 035	71	25	59	2	1 599 000	98	79	63	9
National	24 979	69	25	61	7	11 217 000	97	78	62	28

Note:
a Limpopo was known as Northern Province prior to 2003.

While not strictly a part of the South African context described here, this chapter concludes with a summary of some international comparative studies of science assessment systems, some of which implicitly referred to validity in understanding assessments. This international perspective is given here because these studies have shaped the methodology of this study. For example, a within-country study of an assessment system, such as this study of South African SC examinations, could miss invariant features which might be important in understanding student learning. Comparative studies between countries could "throw into sharp relief the invariant and taken-for-granted features of an education system" (Little, 1996, p. 7). Similarly, Drake (1996) noted that when the comparability of assessments between countries was found to be low, the comparison of their assessments could challenge prevailing assumptions and promote a better understanding of problems and possible solutions inherent in testing systems. The studies reported here each used different methods in their research and illustrate that historically different research purposes, and different perspectives about educational standards, have defined the contrasting methods used in various studies and the particular variables considered important in each study. In this thesis appropriate, defensible, parts of methodologies from some of the described international studies have been used (Chapter 5), to operationalize the conceptual framework designed to explicate standards in this study. None of the international studies described here used standards or validity as lenses to make comparisons of assessment question papers nor did they analyze candidates' answer scripts. A review of the literature shows that conceptually, standards elicited from assessments can only be substantiated through analyses of **both** assessments and student performances therein. Therefore a conceptual framework to analyze the SC Biology examination in terms of standards was required, since none existed (Chapter 3).

2.1 Processes, events and procedures associated with the examination of Senior Certificate Biology in South Africa (1994 to 2007)

Secondary school leaving examination systems vary greatly between countries (e.g., Britton, Dossey, Eubanks, Eubanks, Gisselberg, Hawkins, Raizen & Tamir, 1996a; Britton, Hawkins & Gandal, 1996b; Eckstein & Noah, 1993; Umalusi, 2007, Valverde & Schmidt, 2000).. This section describes changes in the South African SC examination processes during the 1994 to 2007 period covered by this study and outlines in particular how these changes affected the examination of Biology. Included is a description of the policy documents that determined the South African Biology syllabus and how Biology was thus examined at SC level.

The author did not set out to report all the contextual details about the SC Biology examinations which appear in this chapter. However, it became clear as she started looking for records about the teaching and learning of Biology in South African high schools, in particular those pertaining to the SC Biology examinations, that such historical contextual information was no longer in the public domain.[9] Consequently the author increasingly had to draw on her own records and some from the private holdings of others who had worked in education over the years in South Africa. This chapter captures as completely as was possible, the context in which the SC Biology examinations were written in the period 1994 to 2007.

2.1.1 The South African Senior Certificate

The SC examinations serve a dual purpose, [10] that is, to serve as both a terminal evaluation of high school as well as a prognostic test for university admission. At the end of Grade 12 (i.e., the SC year) students were awarded a set of marks which indicated their achievement in separate examinations in a combination of a minimum of six subjects of their choice, written at either HG or SG (Joint Matriculation Board [JMB], 1989; DoE, 1997b).[11, 12] The differences between SC Biology HG and SG are discussed in Section 2.1.4. Depending on the marks obtained in all six subjects, whether each subject was written at HG or SG, and the combination of different subjects and levels taken, the SC would have been awarded with or without Matriculation Endorsement (ME) (JMB, 1989; DoE,1997b).[13,14] A high stakes examination, like the SC, was one whose

9 Assessment processes, including examinations, in South African have been described as secretive (Vandeyar & Killen, 2003). As a result of that secrecy, several problems arose while the author was collecting information to write this section of the thesis. Some of the official public education statistics were difficult to source, especially those for the period 1994 and 1995, and discrepancies exist between data sets from different sources. Some documents had different titles on the title page and on the first page of essentially the same document; some documents had obviously incorrect dates; some documents with the same kinds of information had slightly different titles between years and the Biology SC examination question papers had different title pages between years. Wherever possible these discrepancies have been noted where appropriate in the text, and in the reference section, for the reader who is unfamiliar with these idiosyncrasies and to facilitate future research.

10 No single assessment can serve multiple purposes (e.g., National Research Council [NRC], 2006), nor should it be expected to serve multiple purposes (e.g., American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999).

11 Each year new examinations were set and then written in October/November of that year. Separate supplementary examinations were set for later writing in March/April of the following year to cater for students who for acceptable reasons, such as illness, could not write the examination or for students who qualified to rewrite particular subjects.

12 A student's achievement is the sum or aggregation of the marks obtained in all the questions of an examination paper. Candidates can reach the same mark through different routes if they have different strengths and weaknesses. That is, candidates can compensate for their weaknesses by gaining marks for their strengths.

13 Matriculation endorsement has also become known as 'matriculation exemption' in South Africa.

consequences matter (Brennan, 2006a). Employers used SC marks to select potential employees (Reddy, 2006a). Moreover, a SC with a ME certified its holder eligible to proceed to tertiary studies but did not guarantee the holder access to all tertiary institutions. Some institutions were oversubscribed and/or had their own admission criteria and selection.

For candidates who achieved certification in a subject (i.e., a pass), the mark (score) they were awarded in each subject appeared on their SC document which signified a successful completion of high or secondary school. Prior to 1996, the marks for each subject on the SC were reflected as a symbols, namely A to H including FF and GG, denoting different levels of a continuum of student achievement,[15] and thereafter they were reflected as percentages (Table 2.2). Despite a change to reporting as percentages, in the analyses of results, researchers and the media continued to talk about SC results in terms of the symbols. For each of the two levels, HG and SG, the *cut-scores* which separated the symbols A to H *remained the same between years and were the same for all subjects*. The cut-scores for each of the symbols represented the same percentage of the total marks (Table 2.2). However, because HG and SG examinations each carried different total marks, two different marking scales existed (Table 2.2). That is, symbols A to H represented different bands of total marks each with the same labels (A to H) for both HG and SG. In addition, specific categories (symbols) of HG failed outcomes were converted to SG passes (Table 2.2).

The implications of having studied a subject on the HG or on the SG are interesting because they were not always understood. For example, Gilmour and Soudien (2009) stated that a student needed to have taken a subject on the HG in order to qualify for a ME. This claim was not correct − provided a student had sufficient 'other' subjects on the HG they could elect to do 'another' subject on the SG and still qualify for a ME, if their particular combination of subjects and Grades satisfied the ME requirements. Specifically, students writing a non-language subject on HG or SG, could have potentially qualified for a ME in the period of this study. The advantage to students

---

14 For SC without ME, some SG failures were converted to LG passes.

15 Symbols are similar to performance level labels (PLLs) described by Cizek and Bunch (2007) because they label performance categories but they do not inherently carry a description such as 'pass', 'fail', ' basic', 'proficient' and 'advanced'. In South Africa no use is made of performance level descriptions (PLDs) which provide a "fuller, more precise explanation of what the one-word PLLs attempt to convey" (Cizek & Bunch, 2007, p. 46). Within South Africa there is an understanding that an 'A' symbol represents a higher score than a 'B' symbol.

who studied more that the minimum number of HG subjects required for ME was in the calculation of their aggregate mark for the entire SC examination.[16]

Table 2.2 SC HG and SG examination mark ranges and their equivalent symbols (JMB, 1989; DoE, 1997b).

Symbol	%	Mark range HG	Mark range SG
A	80 – 100	320 – 400	240 – 300
B	70 – 79	280 – 319 [a]	210 – 239 [a]
C	60 – 69	240 – 279	180 – 209
D	50 – 59	200 – 239	150 – 179
E	40 – 49	160 – 199 [b]	120 – 149
F	33.3 – 39	133 – 159 [c]	100 – 119 [d]
FF	30 – 33.3	120 – 132 [e]	90 – 99 [f]
G	25 – 29	100 – 119 [g]	75 – 89 [f]
GG	20 – 24	80 – 99	60 – 74
H	0 – 19	0 – 79	0 – 59

Note:

a Result was condoned upwards by a maximum of 2% if candidate could then obtain 80% (DoE, 1997b).

b Minimum pass on HG.

c HG mark between 133 and 159 converted to SG ‘E’ pass (Department of National education [DNE], 1989) for SC with and without ME (Umalusi, 2005).

d Minimum pass on SG.

e HG mark between 120 and 132 converted to SG ‘F’ pass (DNE, 1989) for SC with and without ME (Umalusi, 2005).

f SG passes between 75 and 99 converted to LG ‘F’ pass (DNE, 1989) for SC without ME (Umalusi, 2005). This process was to have been curtailed at the end of 1999 (DoE, 1997b) but was continued until 2007 (Umalusi, 2005).

g HG mark between 100 and 119 converted to SG ‘F’ pass (DNE, 1989) for SC without ME (Umalusi, 2005).

Various other assumptions were made about the relationship between student performance at different levels of achievement on the HG and SG levels. For example, education authorities converted some levels of failure on the HG to a pass on the SG (Table 2.2). This practice assumed an ‘equivalence’ between particular pairs of symbols on HG and SG. Another example was the assumptions made when some universities calculated points for admission into study, that symbols within and across subjects could be scored by two common ranges of points, one for HG and one for SG symbols (Table 2.3). This practice of those universities also assumed an extended

[16] Students who wrote all their non-language subjects on the HG would have had the opportunity to score 200 marks (100 marks per subject) more than a student who elected to do two non-language subjects on the SG.

'equivalence' between particular symbols achieved in examinations written on the HG the symbols achieved in examinations written on the SG.[17] Because this study tests this equivalence between HG and SG symbols (see Chapter 1, research sub-question 3), the relationship between the mark allocation in HG and in SG is shown in the form of a diagram (Figure 2.2) to orientate the reader appropriately.

Table 2.3 The point system used to rank SC marks by some South African universities for admission purposes.

Symbol	HG	SG
A	8 [a]	6 [a]
B	7 [a]	5 [a]
C	6 [a]	4 [a]
D	5 [a]	3 [a]
E	4 [a]	2 [a]
F	3 [a, b]	1 [a]

Note:

a University of Cape Town (2001, 2002, 2003, 2004, 2005, 2006), Science Faculty University of Witwatersrand (Jackson & Young, 1987).

b Two points ī Faculty of Education at the University of Durban-Westville (Behr, 1985).

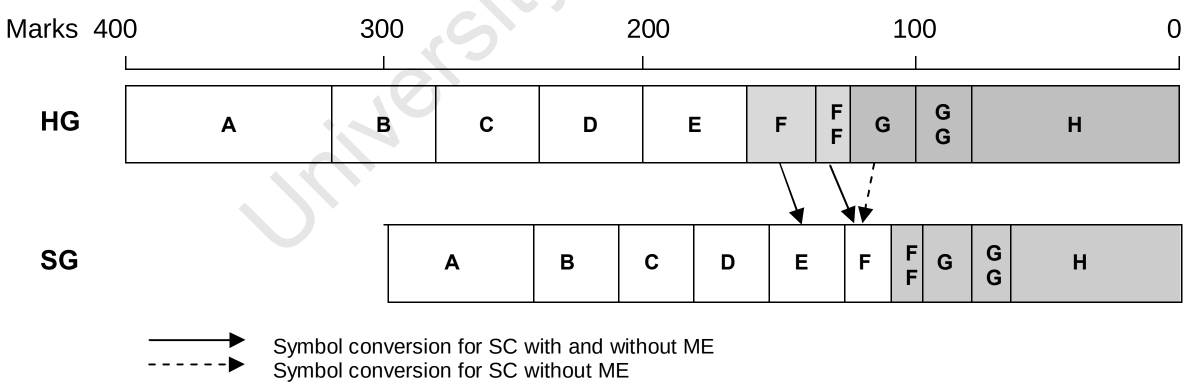


Figure 2.2 A comparison of HG and SG marks and symbols. Shaded areas indicate symbols which denote failure.

[17] For example, this practiced assumed that a 'C' obtained for HG Biology was in some way exchangeable for a 'C' obtained for HG Geography; that 'C' obtained for HG Biology was in some way exchangeable for an 'A' obtained for SG Biology and that an 'C' for HG Biology in two different years was somehow the same.

2.1.2 The South African Senior Certificate examining bodies

During the 1994 to 2007 period of this study, different combinations of SC examining bodies existed in South Africa. Different examining bodies also had different responsibilities in the SC examination process of the country. Each examining body was responsible for setting and/or administering and/or marking SC examinations.

Prior to 1994, the state education and training process was separated largely (exceptions include some non-government schools) on ethnic and racial lines based on the ideologies of the government of the time into a number of different education departments.[18] Broadly, education for white[19] students was governed by the House of Assembly (HOA) and provincial departments, for coloured[13] students by the House of Representatives (HOR), for Indian[13] students by the House of Delegates (HOD) and for black[13] students, the Department of Education and Training (DET).[20, 21, 22] While all these departments fell under the jurisdiction of the Department of National Education (DNE), there was a notional autonomy especially in the flexibility that different examining bodies had with regard to the curriculum (NEPI, 1992a) and each had its own examining body (Table 2.4). In 1996, a single national DoE and nine provincial education departments, each also constituting an examining body, began operating as a result of the changed government of 1994. However, until the end of 1995, the government SC examinations continued to be set, written and administered by the examining bodies recognised by the education sub-departments prior to 1994 (Table 2.4).

---

18 The number of education departments reported at different times varies between different authors depending on their interpretations of the systems of education governance of the ten 'homeland' states - the six self-governing territories administered by the Department of National Education and the four Transkei, Bophuthatswana, Venda, and Ciskei (TBVC) territories recognized by the apartheid government (NEPI, 1992a).

19 The terms 'white', 'coloured', 'Indian' and 'black' were defined by apartheid ideology to distinguish between populations of people of South Africa.

20 Nine of the 'homeland' states followed the DET curriculum at SC level (NEPI, 1992a).

21 Schools designated for black, coloured or Indian students were considered as 'disadvantaged' schools and those for white students as 'advantaged' schools, in part, because prior to 1994 public schools for white students were better funded (Bhorat & Oosthuizen, 2006). Even though increased spending toward poorer schools has taken place in the post-1994 period (DoE, 2003a) the terms 'advantaged' and 'disadvantaged' continue to be used to differentiate between schools because the legacy of racial inequalities pre-1994 persist today (Bhorat & Oosthuizen, 2006; Fleisch, 2008).

22 In 1976 private schools began accepting students outside of the racial groups for which they were authorized by the state to serve (Boner, 2000) and the Interim Constitution Act Number 200 of 1993 that repealed all the apartheid laws ensured that by 1994 all public schools were open to all (Maile, 2004; Selod & Zenou, 2003).

Table 2.4 Examining bodies responsible for the setting and administration of SC Biology examinations from 1994 to 2007.

1994, 1995	1996 to 2000	2001 to 2007
Government	*Government*	*Government*
House of Assembly [a]	Western Cape [h]	Department of Education [k]
Cape [b]	Free State [h]	
Orange Free State [b]	Gauteng [h]	*Non-government*
Transvaal [b]	KwaZulu-Natal [h]	Independent Examinations Board
Natal [b]	Eastern Cape [h]	
House of Representatives [c]	Northern Cape [h]	Beweging vir Christelik-Volkseie Onderwys [l]
House of Delegates [d]	North West [h]	
Department of Education & Training [e]	Northern Province / Limpopo [h, i]	
	Mpumalanga [h]	
National Examination Board [f]		
	Non-government	
Non-government	Independent Examinations Board	
Independent Examinations Board [g]	School of Tomorrow [j]	

Note:

a Administered education for ówhiteô students. In 1991, desegregation in state schools was permitted under certain conditions (NEPI, 1992a).

b Department of Education and Culture Provincial Education Departments set and administered examinations for ówhiteô candidates at state schools within each province.

c Department of Education and Culture set and administered examinations for ócolouredô candidates at state schools throughout South Africa. In 1991 desegregation was permitted under certain conditions (NEPI, 1992a). Also known as HOR.

d Department of Education and Culture set and administered examinations for óIndianô candidates at state schools throughout South Africa. In 1991 desegregation was permitted under certain conditions (NEPI, 1992a). Also known as HOD.

e This examining body set and administered examinations for óblackô candidates at state schools throughout South Africa and included those candidates from schools in some of the independent homeland states. Also known as DET.

f Department of National Education set and administered examinations for candidates at some independent (private) schools and colleges not registered with a provincial education department or with the Independent Examinations Board. Also known as NEB.

g Set and administered examinations for candidates at private schools registered with the Independent Examinations Board. Also known as IEB.

h Provincial Education Departments within each of the provinces set and administered examinations for all candidates from state schools registered within their provinces.

i In 2003 Northern Province was renamed Limpopo.

j Set and administered examinations for candidates at some private schools registered with the School of Tomorrow (SoT) for a few years (exact years could not be determined). The SoT were not part of this study because question papers were not available.

k Set examinations in some subjects, including Biology for all candidates registered at state schools in all provinces. The provinces administered the writing and marking of the examinations. Also known as DoE.

l Administered examinations for candidates at some private schools registered with the Beweging vir Christelik-Volkseie Onderwys (BCVO). The BCVO were not part of this study because question papers were not set and written in English.

From 1996 until 2000, the SC examinations for government schools became the responsibility of the newly recognized, non-ethnic and non-racial provincial sub-departments of the DoE (Table 2.4). The equivalence between the SC syllabuses used and the SC examinations set by the different examining bodies that was assumed or projected by the state,[23] or used for selection purposes by universities (Herman, 1995), became increasing challenged (Muller, 2004). This resulted in a decision in 2001 to pilot a single SC examination for the whole country[24] in each of five subjects, English Second Language, Biology, Physical Science, Mathematics and Accounting, in preparation for the introduction of a new NSC examination to be written for the first time in 2008. In 2004 an additional subject, History, was included and in 2006 five more subjects, Geography, Economics, Business Economics, Agricultural Science and Afrikaans Second Language, were brought to the national examination list. The single national DoE SC examinations for each subject were set by the DoE and administered and marked by the provincial examining bodies.

The Independent Examinations Board (IEB) remained a SC examining body servicing some independent schools through the period of this study (Table 2.4) by setting, administering and marking their SC examinations. Two other examining bodies, namely, the School of Tomorrow (SoT) and the Beweging vir Christelik-Volkseie Onderwys (BCVO) were created in the period 1997 to 2007 but are not included in this research because comparable SC examination question papers could not be obtained (SoT) and because the BCVO examinations were only set in Afrikaans. Both the SoT and the BCVO had fewer students than the other examining bodies.

2.1.3 The South African Senior Certificate examination cycle

The SC examination is the product of partnerships defined for the period 1996 to 2007 in the *National Education Policy Act No. 27 of 1996* (DoE, 1996), the *National Policy on the Conduct, Administration and Management of the Assessment for the Senior Certificate* (DoE, 2004a) and the *General and Further Education and Training Quality Assurance Act No 58 of 2001* (DoE, 2001b). The 'partners' in the process were, with varying responsibilities over the time period of this study, the Minister of Education, the central government (DoE), the provinces (provincial education departments), the independent or private examining bodies and a succession of two certification

23 See Christie (1987) for claims made by state officials that the syllabuses and examinations of the DET for black students were the same as those used for non-black students.

24 Students in government schools and students in private schools registered with each of the provincial education departments.

bodies, the South African Certification Council (Safcert) from 1994 to 2001 and Umalusi from 2002 to 2007. Safcert and Umalusi were responsible, at different times, for the quality control of the SC examinations and the issuing of SCs to successful candidates (Lolwana, 2006). The differing responsibilities of these partners are briefly outlined below.

During the period 1994 to 2000, SC examinations were set each year and administered by the examining bodies listed in Table 2.4, with Safcert performing a standard-keeping function (Lolwana, 2006) so as to ensure the same standard of education and examining (Muller, 2004). This era of Safcert "has not been formally documented" and "accounts [of their practice]… are primarily anecdotal" (Lolwana, 2006, p. 20). Consequently, the author was unable to find documented descriptions of the details related to how each of the examining bodies of the time practiced their SC examinations process. The first documented record of the examination process, named the SC examination cycle, appeared in DoE (200[5]a) and it is this record that has been modified, based on the author's personal experience[25] (Figure 2.3). Figure 2.3 is representative of the process from 2001 until 2007 for national SC examinations set by the DoE.[26] Prior to 2001 there is no recorded evidence of the exact processes followed by each of the separate examining bodies, nor evidence whether the SC examination cycle of each examining body was the same. In descriptions of examination cycles, the relative importance of the different elements in a cycle is dependent on the point of view of the creator or reader of the description and their personal involvement (Williamson, 2003) or interest. Because this study focuses on the analyses of SC examination question papers, the elucidation of the examination cycle will begin with a description of how the examinations are set (Figure 2.3 [*A*]), thereafter followed by a description of the marking review process (Figure 2.3 [*B*]) and how the raw scores are standardized (Figure 2.3 [*C*]). Raw scores refer to the total marks that candidates are allocated by markers who score their examination scripts and which may or may not be statistically adjusted during the standardization of marks meeting prior to the public release of the results.

2.1.3.1 *Setting of question papers*

The process of the setting, and the subsequent internal moderation of SC examination question papers and answer scripts have changed since the 1990s. The examining bodies changed during this

---

25 The author has served as an External Moderators for Biology from 1990 to the present time.

26 The SC examination cycle represented here is similar to that followed by the IEB between 2001 and 2007.

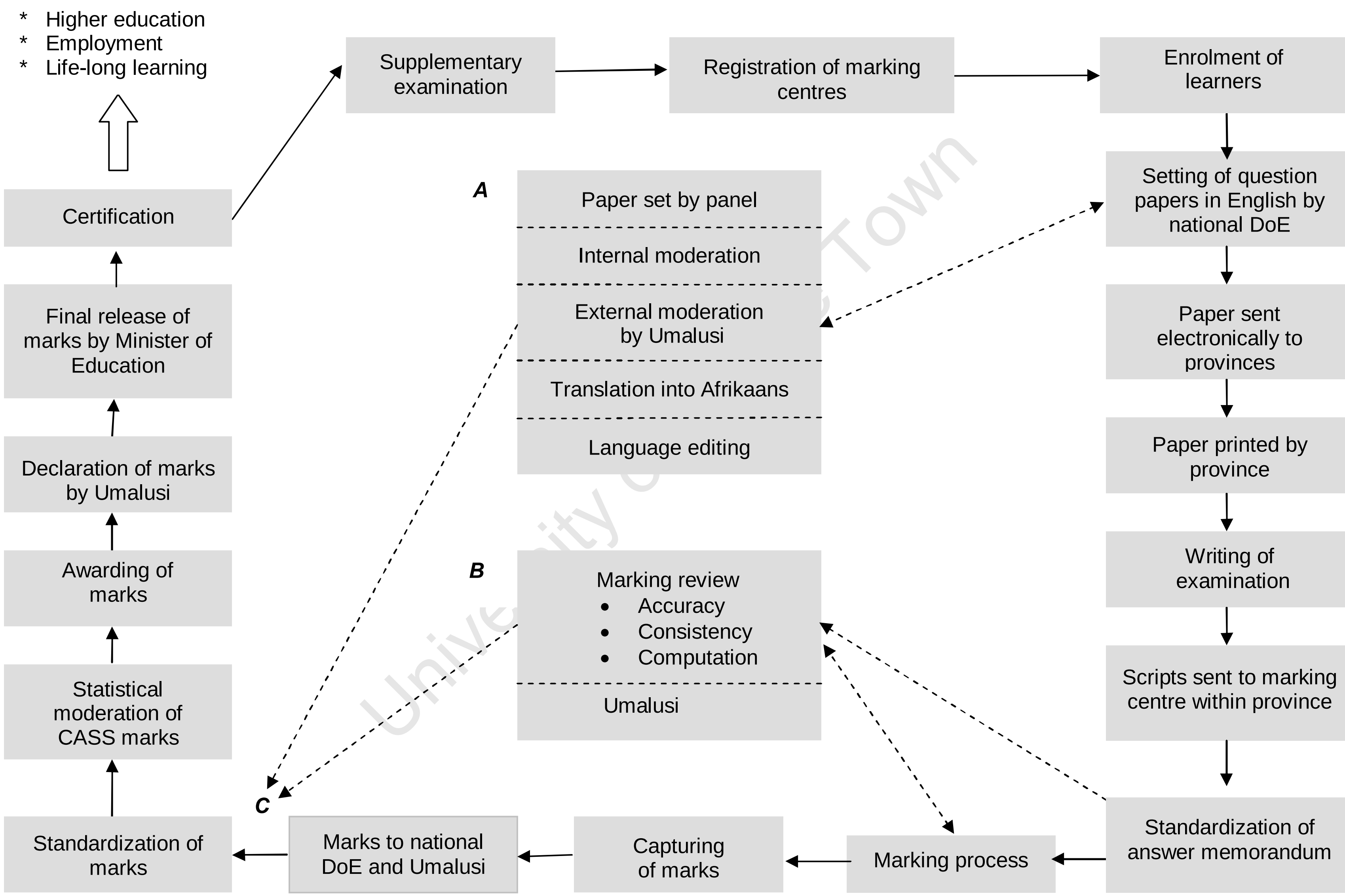


Figure 2.3 The Biology SC examination cycle for state schools (Modified from DoE, 200[5]a).

time (see above): single examiners per paper were replaced by a panel of examiners, and the role of external moderation by the JMB,[27] Safcert[28] or Umalusi, evolved to be overall quality control rather than simply ensuring the integrity of the subject/discipline as it had been in the early 1990s (Lolwana, 2006). Prior to 2001, external moderators received question papers in the mail for one moderation after which it was the responsibility of the examining body to make (or not make) changes suggested by the external moderators. Two external moderators for Biology were replaced in 2005 by a panel of three (2005 and 2006) or four (2007) external moderators who were responsible for moderating both the question papers and the Continuous Assessment (CASS) component[29,30] awarded for formative assessments[31] done by candidates during the year.

In 2001, Biology became a nationally examined subject, that is, all candidates (except for those registered with the non-government examining bodies) wrote the same nationally set examination. That year the single SC examination question paper of previous years was divided into two question papers[32] and, for the first time, national official weighting was given to various components, topics,[33] of the syllabus in the final examinations (Table 2.5). Prior to 2001, some examining bodies (e.g., Cape of Good Hope, 1986; WCED, 1996) had their own suggestions for teachers and examiners as to the relative emphases of the different topics (Table 2.5). These suggestions were not mandatory.

The setting of the question papers (Figure 2.3 [*A*]) to be written at the end of each year[34] began near the start of that year. Between 2001 to 2007, the Biology examinations question papers were set by a panel consisting of between four and eight DoE appointed examiners who were all employed by the provincial education departments (but not as teachers). In the earlier years (i.e., 2001 to 2004), four examiners (including one chief examiner) were assigned to each of the two papers. More recently (i.e., 2005 to 2007), a panel of four examiners (including one chief examiner) became responsible for both the question papers. During the setting stage, question papers move

27 The JMB ceased functioning with respect to the SC examinations in 1992 (Lolwana, 2006).
28 Safcert ceased functioning with respect to the SC examinations in 2001 (Lolwana, 2006).
29 A SC CASS component became compulsory for all students from 2001.
30 More recently known as School Based Assessment (SBA) (Poliah, 2009).
31 Although called formative assessments by policies, many of these assessments were more like the *interim tests* described by Popham (2009) because these assignments were tests similar in structure and content to the SC examination.
32 The IEB continued to write one examination question paper.
33 In this context ‘topic’ is used to refer to a group of facts, concepts and processes which are grouped according to some organizing principle recognized in Biology. In official DoE documents these elements are called ‘content’.
34 Supplementary questions papers, to be written in the first quarter of the following year, were also set for students who for legitimate reasons could not write the end-of-year examinations.

Table 2.5 Examples of changes in emphasis on different topics in SC during the period 1994 to 2007.

Topic	**CED** (% teaching time) [a]	**IEB** (% teaching time) [b]	**House of Delegates** (% weighting in examination) [c]	**WCED** (% weighting in examination) [d]	**KwaZulu-Natal** (% weighting in examination) [e]	**DoE** (% weighting in examination) [f, g]
Biological Compounds	15	11	10	8	7	9
Enzymes & co- enzymes	3			2		
Angiosperm physiology	20		18	16		
Water relations		11			11	15
Growth & development					4	2
Photosynthesis		6			10	7
Cellular respiration	5	6	3	6	6	6
Aspects of human anatomy and physiology			55 [h, i]	46 [h]		
Nutrition	10	11			11	11 [h]
Gaseous exchange	6	11			8	8 [h]
Excretion [i]	6	11			9	11 [h]
Co-ordination	18	18			14	17 [h]
Blood vascular system		10				
Homeostasis [j]	7	6		10	8	5
Population dynamics	10		14	12	12	9

Note:

a Cape Education Department (CED), calculated from teaching time, but no stipulation as to emphases in the SC examination (Provincial Administration of the Cape of Good Hope, 1987).
b Calculated from teaching time, but no stipulation as to emphases in the SC examination (IEB, 1996).
c Isaac (1990).
d WCED (1996).
e KwaZulu-Natal Education Department (1996).
f DoE (2001a).
g Examination split into two papers.
h Includes aspects of homeostasis.
i Includes osmoregulation in *Amoeba* and *Lumbricus* until end of 2002.
j Thermoregulation and tissue fluid.

between the examiners and an internal moderator (also appointed by the DoE and also a provincial education department employee) as they are revised. Moderation of the question papers by external moderators appointed by Umalusi occurred as often as was necessary and in consultation with the panel of examiners and the internal moderator.[35] Once the question papers were approved by the Umalusi external moderators, they were translated into Afrikaans, after which both the English and Afrikaans versions were sent to non-subject specialists for language editing. The international best practice of back-translation of the same question paper appearing in two or more different languages (Abedi, 2006), and the use of pre-testing (Cizek, 2006) did not take place. This substandard practice led to some questions having to be declared as invalid after the writing of the question paper in the two years, 2005 and 2006, for which candidates' answer scripts were analyzed in this study (see Chapter 5).

From 2001 the approved examination question papers were then distributed to each of the provinces where they were printed (Figure 2.3). Each of the provinces then administered the nationally set examinations, and was responsible for the marking of students' scripts. Each examination was a standardized examination as all students taking the examination wrote the same question paper, at the same time, on the same day, under the same administrative conditions. Before the answer scripts were marked, meetings were held at which every province was represented together with the examiners, the internal moderator and the external moderators. The meetings were held in order to standardize the memorandum[36] for each question paper and to establish the principles of marking (Appendix 2.1). The rules of marking and the answers agreed to for each question were then used as a strategy to ensure consistent marking within each province. The discussions from these meetings fed into the marking review process discussed in the next section. Each answer script was hand-scored by markers appointed by each province or examining body. Over the years, delegates at memorandum meetings agreed in principle to not have any one script marked by only one marker and post-2000 this condition has been checked by Umalusi. Pre-2001 there is little evidence available to substantiate how widely multiple question marking per Biology script was practiced by the different examining bodies. In addition, each examining body had internal moderation processes in place through which marking was checked for consistency and accuracy. Reports from the internal moderation processes also fed into the marking review process discussed in the next section.

[35] Prior to 2001, external moderators received the question paper once (in the mail), or twice if the paper was rejected, and had no direct contact with examiners. Sometimes telephonic contact occurred.

[36] A scoring guide is called memorandum in South Africa. The reasons for this terminology could not be traced.

2.1.3.2 *Marking review*

The marking review process (Figure 2.3 [*B*]) occurred on two levels, namely, within the provincial marking centres where the answer scripts were marked and then later reviewed by Umalusi. The accuracy and consistency with which the markers were grading the scripts (with respect to the standardized memorandum), and the accuracy of their computations were checked within (by the provinces) and between provinces (by Umalusi). In 2005, Umalusi conducted a pilot project using Biology scripts which ensured that comments about the accuracy and consistency of the marking process were immediately reported back to each province while they were still marking. In subsequent years, this Umalusi process became the normal practice for many of the SC subjects.[37] Once the candidates answer scripts were marked, the marks were captured electronically under different protocols at dispersed venues and sent to the national DoE and Umalusi for the standardization of the marks (Figure 2.3 [*C*]).

2.1.3.3 *Standardization of marks*

The SC examination is considered "a general norm-referenced school-leaving standard" (NEPI, 1992b, p.136). 'Norm-referenced' in the South African practice means something subtly different to the way the term norm-referenced is used elsewhere. In other countries, such as, for example, the USA, norm-referenced standards are a function of the performance of test-takers in relation to one another (Cizek & Bunch, 2007) and all who have taken the same test,[38] or some comparison group (Linn, 2006). Tognolini and Stanley (2007, p.130) indicated that "[o]ne of the main advantages of norm referencing is that the marks, grades or awards are interpreted in the same way from situation to situation (year-to-year; subject-to-subject)". This inference could only be true if the same norm group was used every year. It is unclear how the same norm group could be used for different subjects. Characteristics of the norm group can change between years and may be different between subjects and will depend on characteristics of each examination with which they were normed. In South Africa, since 1918 the SC raw scores have been statistically adjusted by the JMB, Safcert and Umalusi (Rakometsi, 2011) to "achieve equivalence of the standard of the S[enior] C[ertificate] E[xamination]" across: years, subjects and examination bodies (Fatti, 2006, p.46) (Figure 2.3 [*C*]). The statistical adjustment, for each subject, and for each of the levels HG and SG, involved a comparison of the distribution a current year's total marks (as raw scores) with distributions of previous years by computing a norm from the raw scores of the previous three to

37 The potential benefit of this feedback process to the marking of scripts was unfortunately undermined because marking was not synchronized between the provinces.

38 Professor Saul Geiser, personal communication, February 24, 2010.

five years (Fatti, 2006). The rank order of candidates remained unchanged. This form of statistical adjustment assumed that "the standards of the examinations, including the examination papers, the memorandums and the marking were satisfactory during those previous five years" (DoE, 1998, p. 25). In South Africa, different examination question papers were set each year and for Biology there was no direct overlap between questions asked between years. This non-comparability leads one to question the extent to which the question papers on which the statistical adjustments were made really were equivalent as assumed, especially given the "highly 'political' [nature of the] standardi[z]ation meeting" (Fatti, 2006, p. 49).[39] The process of external moderation and standardization of marks that was done by Umalusi was considered an effort to address the equivalence of question papers between examining bodies (wa Kivilu, 2006). This study seeks to address the issue of equivalence for Biology SC examinations (see Chapter 1, research sub-questions 4 and 7), especially since "comparability of [examination] standards does not mean identity of performance" (Nuttall, 1979, p. 16).

Historically, the different examining bodies used the written SC Biology examination component in different ways, when calculating the final mark that appeared on the SC during the period of this study. Prior to 2001, the policy was flexible enough to allow individual examining bodies to include (or not include) a year-mark[40] or a practical mark in their final mark. For example, in 1994 and 1995, in all but four of the examining bodies, the marks for only the written examinations were used to calculate the final mark. In some cases—for example, the now defunct Natal, Transvaal and Orange Free State examining bodies (Fatti, 2006; Kanjee, 2006) and the IEB—a school-based year-mark and/or practical examination mark was generated, moderated by subject inspectors and used in calculating the final mark. The writing of the first examinations under the examining bodies of the new government, in 1996, resulted in the written examination mark being used together with the teacher generated year-mark, to derive a total mark or composite score for the SC. During this time, some of the new examining bodies (e.g., KwaZulu-Natal, Free State) had a separate practical examination mark which formed a part of the teacher-generated mark. The author was unable to determine how, or indeed if, year-marks and practical examination marks

39 Statistical adjustments are often contested by different examining bodies.

40 Some examining bodies required a year-mark to be submitted for each candidate. This year-mark was a summary of a candidate's performance in both formative and summative assessments during the year.

which were included in the SC final mark were verified or statistically adjusted by Safcert during the 1994 to 2000 period.[41]

In 2001, the year-mark became known as the Continuous Assessment or CASS mark, and it became a compulsory component of the scores of all students, including those in independent schools. In Biology, the CASS mark was allocated by individual teachers for syllabus-based work completed during the Grade 12 year, and was moderated by examining bodies in accordance with policy that was regulated by the DoE and Umalusi. Between 1996 and 2007, 25% of the final SC examination mark in each subject was assigned to the CASS component. The composition and weighting of the various components of the CASS mark varied during the period 2001 to 2004 in order to address and allow for the unequal distribution of resources required to do CASS work (DoE, 2002a). From 2005 to 2007, the composition and weighting of CASS tasks were prescribed (DoE, 2005b). Lubisi and Murphy (2002) have documented historical and current problems associated with the implementation and use of the year-marks in South Africa.[42] Many of these problems resulted from teachers not understanding what formative assessment is and how it should be used in the classroom, problems which were similar to assessment problems in other developing countries such as Sri Lanka and Ghana (Lubisi & Murphy, 2002). As a result, the implementation, monitoring and moderation of the CASS component across the country proved problematic and unreliable, and the CASS marks were often much higher than the corresponding marks achieved in the written examination, particularly in the scientific subjects (Fatti, 2006). Because of this lack of confidence in the reliability of CASS scores at the end of the year (Poliah, 2009),[43] the CASS marks of each school were statistically moderated by the NDoE and Umalusi so that they were collectively within 5-10% of the mean marks obtained in the final SC examination (raw scores) by the school concerned (Fatti, 2006).[44] Non-language examination question papers, such as Biology, were set in English and in Afrikaans only. Thereafter candidates who did not have English or

---

41 The author was unable to confirm which of the examining bodies had a practical component as part of their SC examinations. No records containing this information could be located but the author does know, from her personal experience as an external moderator, that no practical examinations were externally moderated during the period of this study.

42 Lubisi and Murphy (2002) referred to school-generated marks as formative assessment marks although marks from summative assessment tasks, like tests and mid-year examinations (interim tests [Popham, 2009]), often contributed to these mark.

43 Amedahe (2001) noted similar concerns about the use of teacher-based continuous assessment scores with external-examination scores for certification in Ghana.

44 Willingham, Pollack and Lewis (2002) proposed a framework of possible differences between measures of student achievement obtained from classroom activities and measures of student achievement obtained from tests, in high-stakes assessment scenarios in the USA. Willingham et al. (2002) offered statistical ways in which the relationship between these two measures of student achievement might be understood so that their complementary strengths might be enhanced.

Afrikaans as their first language[45] were awarded an extra 5 per cent of the mark which they obtained in the examination (Fatti, 2006), equivalent to 1.05 times their own final percentage. The statistically adjusted CASS mark was then combined with the statistically adjusted written examination mark to give the final mark awarded to each candidate, that is, the mark that appeared on the SC.[46]

The standardization processes used in South Africa have been subject to debate and still are.[47] Some of these reasons for debate have been discussed above. Another concern is because the statistical processes were developed assuming a stable population rather than a transforming population and did not take into account changing norms (Lolwana, 2006). The norms, in the South African context, are specifically norms in student performance and are applied as if largely independent of the difficulty of examination question papers, the memoranda and the quality of the marking. Even if the raw scores from different years are statistically equated to maintain a consistent scale (or standard), comparable content across years is required to equate scores (Kolen & Brennan, 1995). In the South African context in recent years, written and verbal comments from examining bodies, examiners and an external moderator[48] about the fairness and difficulty of examination question papers and the results of the marking review process were considered during the statistical moderation process (Fatti, 2006) (Figure 2.3). The implications of some of these practices are addressed where appropriate in this thesis.

2.1.4 The South African Senior Certificate Biology Syllabus

A syllabus known as the Core Biology Syllabus (CBS) (DNE, 1984a, 1984b), stipulated what was to be taught and examined at either the HG level or the SG level in each of the three last years of high school, Grades 10 to 12[49] and examined in the SC Biology examination at the end of Grade 12. Differences between the respective HG and SG for Grade 12 were minimal. For

[45] In South Africa, students for whom English is not a first or home language are currently called English as Additional Language Learners (EALLs) which means the same as English Language Learners (ELLs) in the USA (C. McKinney, personal communication, March 15, 2011).

[46] The CASS mark, its statistical adjustment and the adjustment for language do not form a part of this study. The process is explained here to complete the context of this study.

[47] In February 2011, due to public pressure Umalusi made public the adjustments that were made to marks in the 2010 NSC examinations (Umalusi, 2011). This formal disclosure is now adopted as Umalusi policy.

[48] Since 2006 one Umalusi Biology external moderator has been include in the standardization of Biology SC marks.

[49] Prior to 1996 the three years of school that culminated in the SC examination were known as Standards 8, 9 and 10. From 1996 these same years become known as Grades 10, 11 and 12.

example, HG students needed to learn the functions of different parts of the brain and SG students needed to learn the general functions of the brain. The first SC Biology examination based on the CBS was examined at the end of 1987. The different examining bodies that existed at the time used the CBS to compile their own syllabuses (NEPI, 1992a) and, in some cases, their own versions of guideline documents to guide their teachers. The only policy requirement was that examining bodies could not remove anything from the CBS (NEPI, 1992a). Consequently, different examining bodies emphasized different parts of the CBS (Table 2.5). The different policies summarized in Table 2.6 determined the use of the CBS during the period covered by this study. The different policies are concerned with various interpretations of the CBS and these interpretations are discussed in the following discussions about the CBS.

Table 2.6 Summary of the sources of policy and documents which determined the version of the CBS used in 1994 to 2007.

Source	Title of policy	CBS reference number
DNE (1989)	A résumé of instructional programmes in public ordinary schools Nated 02-550 (89/03)	HG 153101010 [a] SG 153203510 [a]
DoE (1997b)	A résumé of instructional programmes in public schools Report 550 (97/06)	HG 153102212 SG 153202512
DoE (2001d)	A résumé of instructional programmes in schools Report 550 (2001/08)	HG 153102212 SG 153202512
DoE (2001a)	National Senior Certificate Examinations Guideline Document Biology HG & SG Papers 1 & 2 [b]	HG 153102210 SG 153202510
DoE (2002a)	National Senior Certificate Examinations Guideline Document ī Biology HG & SG Papers 1 & 2 (From 2002) [c]	HG 153102210 SG 153202510
DoE (2005b)	A résumé of instructional programmes in schools Report 550 (2005/09)	HG 153102212 SG 153202512

Note:

a These reference numbers do not appear on either of the two copies of the CBS obtained in 2005 from two reputable sources: Mr Cobus Lötter who was the Director: Matriculation Board, Committee of University Principals at the time and Mr Peter Ayerst who served on the committee responsible for drawing up the syllabuses.

b This document became *de facto* policy in 2001 when the first nationally-set Biology SC examination was written for candidates from state schools. Part of this document was concerned with elucidating the CBS referred to in DoE (1997b).

c This document became *de facto* policy from 2002 to 2007 for candidates from state schools. Part of this document was concerned with further elucidating the revised CBS which was done in DoE (2001a).

The CBS (DNE, 1984a, 1984b) provided a general list of learning objectives and approaches to the syllabus at the beginning of the document which were expected to apply **across all** the content. They elaborated the content which students at each of the three years of study were expected to learn and, on which they would be assessed (Figure 2.4). In the CBS 'content' was used to describe facts, figures and processes[50] and the objectives and approaches reference levels of cognitive demand (Figure 2.4).[51] While the CBS stipulated some practical work that was to be done each examining body was free to stipulate the specific practical work that was to be covered by schools under their jurisdiction.

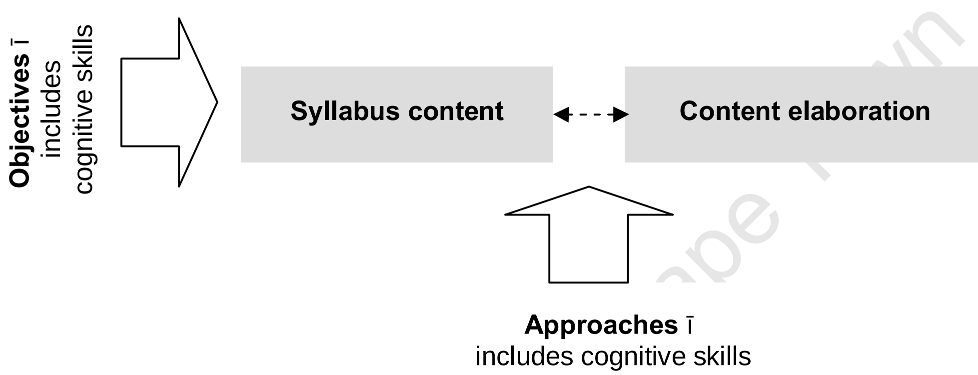


Figure 2.4 The relationships between the objectives and approaches to the content as stipulated in the CBS.

The content in the CBS documents, and in the subsequently produced guideline documents, was presented in what Darling-Hammond (1997) refers to as a laundry-list.[52] That is, the format of the CBS simply included a list of unconnected facts and concepts, grouped as different topics, without highlighting the organizing or main ideas, the relative sequences and/or the relationship between and the essential skills students required to become educated within and between topics. This approach of simply listing topics and subtopics in terms of facts and general skills has been criticized as it gives little direction for the development of applications and conceptual understanding, including higher order thinking and problem solving (Porter, 1989) that "give

50 The CBS uses the term content in a different way to which the same word will be conceptualized in the framework developed in Chapter 3 to guide the analyses of the SC Biology examinations. In Chapter 3, CBS 'content' becomes known as 'topics'.

51 To avoid repetition, the reader is referred to learning objectives and approaches stipulated in the CBS given later in Chapter 4 (Table 4.2 and Table 4.3).

52 An exception was the IEB Guideline Document which presented learning outcomes with each section of content (IEB, 1996).

meaning and utility to facts and skills" (Porter, Archbald & Tytree, 1991, p. 11). The danger of a laundry-list is that the content is often taught and learned in a rote fashion. Also, cognitive "skills have no meaning by themselves" (Marzano & Costa, 1988, p. 69). The CBS laundry list of content, or the syllabus as it is known, is given explicit direction with respect to cognitive demand because the CBS required that all levels of cognitive skills be infused with the entire syllabus through the objectives of the syllabus and through the approaches to the syllabus (Figure 2.4). Therefore the CBS implied that all content should be taught at all cognitive levels and that all content could potentially be examined at any cognitive level. The findings of this study will determine how successfully the marriage of content and cognitive skills implicit in the CBS was implemented in the SC examinations.

In 1995 in preparation for implementation in 1996, each of the new examining bodies reporting to the newly established national DoE formulated their own interim syllabus, still based on the CBS. The next 'interpretation' of the original core syllabus occurred when a single national SC examination was introduced in 2001 for some subjects, including Biology, under the regulation of the DoE, for most students across the country (with the exception of students of the IEB, the SoT and the BCVO). Biology specialists, mainly Biology examiners from the previous years, representing each of the provincial (government) examining bodies then negotiated a single modified CBS, the National Guideline Document for Biology (NGDB) SC examinations (DoE, 2001a) based on their common understanding of the original core syllabus document (DNE, 1984a, 1984b). The national panel of examiners then updated the NGDB in 2002 (DoE, 2002a).[53] The BCVO followed the NGDB and the IEB retained their pre-2000 interpretation of CBS content. Subsequently, additional documents detailing the format of the national Biology SC examination papers (DoE, 2004b), examples of rubrics for marking (DoE, 2004c), and principles of marking (DoE, 2001c, 2002b, 2003a, 2004d, 2005c, 2006a) were developed for public use.

An example of the subtle kinds of changes made to the syllabus by examining bodies in 1987, 1996 and 2001 is shown in Figure 2.5 using the HG core syllabus entry on 'vitamins'.[54] Another change was that while the original CBS (DNE, 1984a, 1984b) did not require that structural formulae for the organic compounds studied be learned, the Transvaal, Cape, Natal, HOR and the IEB

[53] Choppin (1980, pp. 14-15) strongly discouraged the practice of examiners determining the curriculum – "The curriculum is far too important to be left in such hands … assessment centre[s and examiners should be] the servant[s], not the master of the curriculum".

[54] In this figure the author chose to retain the exact format and wording of the original documents − some of which are difficult to locate in the public domain. Therefore, interested readers can get a sense of how these documents looked.

A. CBS (DNE, 1984a, p 21)

1.3.4 Vitamins Biological importance of vitamins	Vitamins as being essential for many metabolic activities; special reference to the roles of some vitamins; their sources in normal human diet, and the consequences of their deficiency.

B. CED interpretation of the CBS (Provincial Administration of the Cape of Good Hope,1987, p. 37, bold emphases the author's own)

1.3.4 Vitamins Biological importance of vitamins	Vitamins as being essential for many metabolic activities; special reference to the roles of **the following** vitamins: **retinol (Vit. A), chole-calciferol (Vit. D), thiamine (Vit. B1)[,] nicotinic acid (Vit. PP), ascorbic acid (Vit. C)**; their sources in normal human diet, and the consequences of their deficiency.

C. HOR interpretation of the CBS (House of Representatives Department of Education, 1986, p. 41, bold emphases the author's own)

1.3.5 Vitamins Biological importance of vitamins	Vitamins as being essential for many metabolic activities; special reference to the roles of **the following** vitamins: **A**, **B**, **C** and **D**; their sources in normal human diet, and the consequences of their deficiency.

D. WCED interpretation of the CBS (WCED, 1995a, p. 29, bold emphases the author's own)

1.3.4 Vitamins (a) Biological importance of vitamins	Vitamins as being essential for many metabolic activities; special reference to the roles of **the following** vitamins; **retinol (Vit. A), cholecalciferol (Vit. D3), thiamine (Vit. B1), ascorbic acid (Vit. C)**; their sources in normal human diet, and the consequences of their deficiency.

E. KwaZulu-Natal Education Department interpretation of the CBS (KwaZulu-Natal Education Department, 1996, p. 3, bold emphases the author's own)

1.3.4 Vitamins (B)iological importance of vitamins	Vitamins as being essential for many metabolic activities. Special reference to the roles of: **Vitamin A, Niacin, Vitamin C and Vitamin D.** Their sources in normal human diet, and the consequences of their deficiency. ([D]eficiency diseases and symptoms)

F. DoE interpretation of the CBS (DoE, 2001a, p. 2, p. 7; DoE, 2002a, p. 6, p. 17)

1.3.4 Vitamins: Biological importance of vitamins

Vitamins as being essential for many metabolic activities; special reference to the roles of Vitamins A; B1; B2; B3; C; D; E and K; their sources in normal human diet, and the consequences of their deficiency.

G. IEB interpretation of the CBS (IEB, 1996, p. 1)

Candidates should be able to:

(n) [L]ist an example of a vitamin that is a co-enzyme and describe the results of its deficiency.

Figure 2.5 Evolution of interpretations of the CBS HG entry for vitamins.

examining bodies required their students to learn the formulae. In the 1996 revision, Northern Province, North West Province, Northern Cape Province, Eastern Cape Province, Gauteng Province and the IEB examining bodies required structural formulae. No structural formulae were required in the 2001 syllabus revision for government schools, but they remained a requirement for IEB students. Pre-1996, the National Examination Board and the IEB required that students learned about C_3 and C_4 plants and the compensation point when learning about photosynthesis and cellular respiration. Pre-1996 some examining bodies required calculations of water potential in plants.

Post-1995 modifications to the CBS appeared to have not been major with respect to what the CBS termed content. Jansen (1999) investigated the post-apartheid changes that were made to the school curriculum. He indicated that the intention of the changes were to "purge the apartheid curriculum of its most offensive racial content and outdated, inaccurate subject matter" (Jansen, 1999, p. 57) and concludes that because of political constraints the endeavour was unsuccessful. In addition, Jansen (1999, p. 63) said that "changed syllabuses simply reflected the existing House of Assembly (that is, the white political establishment for education) syllabuses". Examining the syllabus (Figure 2.5) one might be tempted to endorse Jansen's statement with respect to changes in the 1995 version on the basis of the entry for vitamins resembling as they do what was in the old 'white' Cape Education Department (CED) syllabus. Without access to the complete syllabus documents[55] from 'other' examining bodies it is difficult to say if the statement is correct. It may also be simply that because Biology is a natural science, it is without the inherent socio-political biases that subjects such as, for example, some languages and History may have had. Some of the differences in how examining bodies interpreted the CBS during this period of transformation will become evident especially from analyses of the examination question papers in Chapter 6 and will be reported where appropriate.

The content of curricula, such as syllabuses, should be determined by disciplinary and pedagogical concerns (Schafer, 2011). There was no documentation available explaining the selection of particular topics, and there was no consistent indication of the level of detail required within each of the tropics, for any of the Grades 10 to 12. The choice of content and the lack of a content framework to indicate connections between the topics in the CBS, will not be discussed in this thesis, due to space constraints. For a global perspective, and to orientate readers unfamiliar with

[55] The author tried to obtain these documents but was unsuccessful – copies appear to not have been retained within the national DoE or they cannot be located within the national DoE

South African Biology syllabuses, the topics included in the South African CBS have been mapped on the corresponding augmented Third International Mathematics and Science Study (TIMSS) Curriculum Framework (Britton & Raizen, 1996) (Table 2.7). In South Africa, only the content of Grade 12 was examined in the SC Biology examinations although concepts from Grades 10 and 11 could have been used by students when answering questions (DNE, 1984a, 1984b). The most obvious difference between the topics in the three years of schooling guided by the CBS and the TIMMSS framework were the absence of the topics molecular and population genetics, animal behaviour and evolution from the CBS.

What all the South African syllabus policy documents lacked was the specific **combination** of content (topics) and cognitive skills that Grade 12 (SC) Biology students were expected to master. If these expectations were expressed as learning outcomes or standards, the content of any particular examination could be compared to the expectations and inferences made about how much or how little that examination addressed the required standards. Part of this study develops a conceptual framework (see Chapter 3) and a methodology (see Chapters 4 and 5) to generate *post hoc* content standards as a combination of topics and cognitive skills, in a way that allows comparison of the examination expectations of different examining bodies within years and between years (see Chapter 6).

2.1.5 The South African Senior Certificate Biology examination requirements (1994 to 2007)

The examination requirements for the SC Biology examinations during the period of this study were given in the original CBS (DNE, 1984a, 1984b). These requirements remained in effect until 2001 when the examination was split into two papers and other specifications were made for the government administered examining bodies (DoE, 2001a).[56] The original examination requirements and the subsequent changes are summarized in Table 2.8. Examples of changes which came into effect in 2001 with the introduction of two separate question papers for the DoE Biology SC examination, and which are pertinent to this study, were the number of question papers; a longer time spent answering question papers; the weighting given to different sections of the syllabuses; a requirement that HG students had to answer an essay type question; the use of Bloom's Taxonomy

[56] These examination requirements were the same as those used by KwaZulu-Natal in the 2000 SC examinations. There were no documented reasons supporting the split into two papers, the topic weightings, or the interpretation of cognitive demand (S. Chetty, personal communication, August 15, 2005).

Table 2.7 Summary of the CBS Grades 10, 11and 12 according to the topics identified in the Augmented TIMSS Biology Curriculum Framework (Britton & Raizen, 1996).

Topic	Grade 10 [a]	Grade 11 [b]	Grade 12 [a, b]
Diversity, organization and structure of living things [c]			
Plants			
Algae		Å	
Fungi		Å	
Mosses		Å	
Fern		Å	
Seed-producing plants		Å	
Animals			
Invertebrates			
Unicellular animals		Å	
Coelenterates		Å	
Worms		Å	
Insects		Å	
Spiders [d]			
Vertebrates			
Fishes		Å	
Amphibians		Å	
Reptiles		Å	
Mammals		Å	
Other organisms			
Microorganisms			
Diversity of microorganisms		Å	
Viruses		Å	
[Bacteria] [e]		Å	
Roles in recycling		Å	
Micro-organisms and [hu]man[s]		Å	
Organs, tissues			
Complementari[t]y between structure and function	Å	Å	Å
Cells			
Cell structure and function	Å		
Types of cells	Å		
Cell reproduction	Å	Å	
Life processes and system enabling functions			
Life processes and systems			
Photosynthesis, energy capture, storage and transfer			Å
Respiration, mitochondria			Å
[Gaseous exchange] [e]		Å	Å
Digestion and excretion		Å	Å
Other energy handling [d]			
Sensing and responding			
Biofeedback and homeostasis			Å
Sensory systems, responses to stimuli	Å		Å
Biochemical processes in cells			
Metabolism, protein synthesis, enzymes		Å[f]	Å[f]
Regulation of cell functions			Å
Cell water relationship[s]			Å
Life spirals, genetic continuity, diversity			
Life cycles			
Life cycle of plants, insects, etc		Å	
Reproduction, aging, death		Å	
Cell division, differentiation, succession		Å	
Reproduction			
Reproduction in seed plants		Å	
Sexual reproduction		Å	
Human reproduction		Å	
Vegetative reproduction		Å	
Variation and inheritance			
Meiosis	Å	Å	
Mendelian genetics		Å	
Molecular genetics [d]			
Population genetics [d]			
Biotechnology and application of genetics		Å	

Table 2.7 continued

Topic	Grade 10 [a]	Grade 11 [b]	Grade 12 [a, b]
Evolution, speciation, diversity			
Variation		Å	
Evidence of evolution [d]			
Mechanisms of evolution: Lamarckism [d]			
Implications of evolution [d]			
Biochemistry of genetics			
Structure of DNA		Å	
Replication of DNA		Å	
Transformation of DNA to RNA		Å	
Mutation, gene expression		Å	
Operon model in bacteria [d]			
Implications for society, genetic engineering		Å	
Interactions of living things			
Biomes and ecosystems			
Tundra and deserts [d]			
Rain forest, wetland, other biomes or ecosystems	Å		
Habitats and niches			
Habitats and biotopes	Å		
Niches, endangered species	Å		
Interdependence of Life			
Food chains webs	Å		
Adaptations to habitat conditions	Å	Å	
Competition amongst organisms	Å		Å
Symbiosis, commensalism, parasitism	Å		
[Hu]manǿs impact on the environment	Å		Å
Animal behavio[u]r			
Territorialism	Å		Å
Social grouping (beehives, herds) [d]			
Mating behavio[u]r and selection [d]			
Migration of birds, fishes, butterflies [d]			
Rearing the young [d]			
Learned behavio[u]r [d]			
Human biology and health			
Nutrition			
Foods, vitamins, minerals etc			Å
Balanced diets			Å
Diseases and health			
Prevention of disease, maintaining good health	Å		Å
Causes of diseases	Å		Å
Remedies	Å		Å
Human biology			
Organ systems, organs, tissues [and functioning] [e]	Å		Å
Cells	Å		Å
Energy handling			Å
Sensing and responding	Å		Å
Life cycle		Å	
Reproduction		Å	
Genetics		Å	
Evolution [d]			
Biochemistry of genetics [structure of DNA and RNA only]		Å	
Interdependence of life	Å		
Human behavio[u]r [d]			
[Hu]manǿs impact on environment	Å		Å

Note:
a Known as Standards 8, 9 and 10 until 2001.
b Only content from Grade 12 is examined in the SC Biology Examination.
c Classification of organisms according to TIMSS, rather than current biological taxonomy.
d Not covered in the South African CBS.
e Beyond augmented TIMSS ï covered in South African CBS.
f Protein synthesis only (Grade 11) and metabolism and enzymes only (Grade 12).

Table 2.8 SC Biology Examination requirements for 1987 to 2007.

Characteristic	**CBS (DNE, 1984a, 1984b)**	**Modified CBS (DoE, 2001a)[a]**
Implemented by	Government examining bodies & IEB	Government examining bodies
Time period implemented	1987 to 2000 (2007 IEB)	2001 to 2007
Number of papers	One	Two
Length of paper(s)	Three hours	Two hours each
Sections of question paper(s)	Sections A, B and C (HG) Sections A and B (SG)	Sections A, B and C (HG) Sections A and B (SG)
Sections A and B	Compulsory	Compulsory
Section C (HG only)	Maximum of two questions, one must be answered	One compulsory question in each question paper
Section A	Short answers	Short, objective-type answers
	Recall and higher intellectual abilities	Recall, comprehension and higher intellectual abilities
	25% (HG), 30 to 33.3% (SG) of total marks	30% (HG), 33% (SG) of total marks
Section B	Three (HG), four (SG) questions which may be subdivided	Three (HG), four (SG) questions which may be subdivided into between two and five sub-questions
	Recall and application of knowledge	Recall, comprehension and higher intellectual abilities
	60% (HG), 66.6 to 70% (SG) of total marks	53% (HG), 67% (SG) of total marks
Section C	Data-response and/or structured essay	Approximately half data response and half a mini-essay on same or related topic
	Select, organize information; set up hypotheses; assess data; express ideas logically and systematically	Recall, comprehension and higher intellectual abilities
	15% (HG) of total marks	17% (HG) of total marks
Weighting of topics	Not given	Given
Proportion of marks allocated to higher order questions	40% (HG) 25% (SG)	40% (HG) 20-25% (SG)
Taxonomy for categorizing higher order questions	None given	An interpretation of Bloom's Taxonomy of Educational Objectives
Use of year-mark or CASS mark in determining final mark	Optional, formative assessments, includes practical work	Compulsory, 25% of final mark, includes practical work
Total marks of examination	Not stipulated	400 (HG), 75% of final mark
		300 (SG), 75% of final mark

Note:
a Did not apply to IEB.

of Educational Objectives (BTEO) (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) to ascertain the proportion of higher order cognitive skills (HOCS) questions; the proportion of HOCS questions for SG students was lowered to between 20% and 25% of the total paper; and a formative assessment component became compulsory. Higher order questions were described as comprised of the categories Application, Analysis, Synthesis and Evaluation of BTEO (DoE, 2001a, 2002a). From 2001, the IEB retained one question paper and HOCS questions comprised the categories Analysis, Synthesis and Evaluation of BTEO. The implications of the different DoE and IEB interpretations of the BTEO are discussed in detail Chapters 4 and 6, and the advantages and disadvantages of one question paper versus two question papers are discussed in Chapter 6.

Neither the CBS, the national DoE nor the IEB provided documentation explaining the required format for the SC Biology examinations.

2.2 Areas of research focus

This section describes the rationale behind the focus of this study and from this discussion arise the delimitations inherent to the study which are the result of unique characteristics of the South African SC examination systems.

2.2.1 Why focus on Biology?

In South Africa, Biology is the subject that continuously has the second highest number (after English Second Language) of candidates in the SC examination each year (Umalusi, 2004). Biology was one of the subjects which was first examined nationally rather than provincially (excluding the IEB, and BCVO which maintained their own examinations during the period of this study). Therefore, relationships between different examining bodies, different policies and potentially different SC examinations can be explored in this study.

Internationally, science education has been linked to economic development (Aikenhead, 1997). Mathematics and science education have been targeted as being important because these disciplines are considered to be vital for social and economic development in the 21st century South Africa (DoE, 2004e). Science and technology were identified as key areas of educational investment in post-apartheid South Africa (Naidoo & Lewin, 1998). In many countries Biology, Physical Science and Earth Science are studied together as one compulsory subject, Science, at secondary school exit level. In South Africa, Physical Science and Biology are considered as two separate, optional subjects during the last three years of secondary school. Some students take neither Physical

Science nor Biology, and Physical Science was studied by far fewer students than Biology (Umalusi, 2004). Even though Laugksch (2001) has shown that Biology might not be the best indicator of scientific literacy amongst South African matriculants, it is the only science subject which most students take at Grade 12 level and is therefore the only external and regular (i.e., yearly) measure of scientific literacy that we have in the current South African schooling system. The number of students studying Biology at SC level increased significantly during the period of this study in both government schools (Table 2.9) and in schools administered by the IEB (Table 2.10).[57] The performance of Biology students during the period of this study was good in the IEB schools (Table 2.10), but very poor in the government schools (Table 2.9) which also educated the majority of students. Many students in government schools who sat the HG examinations did not pass them and some of these students received passes converted to the SG level (Table 2.9). During the period of this study increasingly more students chose to write Biology on the SG rather than on the HG in government schools (Table 2.9). We need to understand the implications of HG versus SG levels of performance in Biology SC examinations for scientific literacy in South Africa in order to address it effectively.

The 21st century has been referred to as the "century of biology" ("Biology's Big Bang", 2007; Kafatos & Eisner, 2004; Meagher & Futuyma 2001; Venter & Cohen, 2004), implying that biological knowledge will become the leading science in its impacts on society in the future. For the next decade or so in South Africa the strongest economic growth will be influenced by progress made in scientific fields such as medicine, biotechnology, and information technology (DoE, 2002c). Therefore, South Africa should actively promote learning and research in these areas now, in order first, to benefit from the growth and jobs that these fields may bring, and second, to promote general understanding by the public about scientific matters. If Biology[58] is the science most favoured by SC students, Biology teaching and assessment will be required to address any consequences that may emerge from society's scientific needs. Policies and interventions aimed at improving and assessing the levels of science teaching and learning in South African schools tends to consider only Physical Science and not Biology (e.g., DoE, 2004e), promoting a misconception that Biology is not as 'scientific' as Physics or Chemistry. Recently it was stated in a widely-read South African newspaper that more students choose to study Life Sciences (Biology) rather Physical Sciences in their last year of school because Life Sciences "requires less maths and will yield higher pass rates (Metcalfe, Orkin & Glennie, 2012, p. 6). Biology teaching and assessment, including examination question papers potentially promote and assess a different combination

57 Some of the figures given in Table 2.9 and Table 2.10 differ from figures given in other sources (e.g. Asmal,1999; Seekings, 2002), indicating inconsistencies in public records, where such records exist.

58 Biology was replaced by Life Sciences in SC examinations from 2008. Biology is the term used in this chapter as it is what the subject of this thesis, even when referring to its role in the future.

Table 2.9 Number of students who wrote the government SC Biology HG and SG examinations 1994 to 2007. Pass rates determined using adjusted marks not raw scores. (− indicates that data could not be sourced).

			Biology HG			Biology SG		
Year	**Number of SC candidates**	**SC overall pass rate**	**Number of candidates**	**Pass HG (%)**	**Pass SG [a] (%)**	**Number of candidates**	**Pass SG (%)**	**Pass LG [b] (%)**
1994	495 408 [c]	58.0 [d]	ī	ī	ī	ī	ī	ī
1995	531 453 [c]	56.0 [d]	ī	ī	ī	ī	ī	ī
1996	518 077 [c]	56.5 [d]	257 614 [e]	32.8 [f]	34.5 [f]	67 238 [e]	46.3 [f]	23.9 [f]
1997	538 189 [c]	47.4 [d]	346 215 [e]	23.8 [f]	26.1 [f]	128 243 [e]	42.5 [f]	22.6 [f]
1998	552 862 [c]	49.3 [d]	287 632 [e]	23.0 [f]	21.9 [f]	163 037 [e]	39.3 [f]	24.5 [f]
1999	511 474 [d]	48.9 [d]	195 004 [d]	22.7 [g]	23.5 [g]	192 783 [d]	38.0 [g]	20.4 [g]
2000	489 941 [d]	57.9 [d]	134 361 [d]	30.0 [g]	25.3 [g]	221 854 [d]	41.6 [g]	21.1 [g]
2001	449 371 [d]	61.7 [d]	106 322 [d]	37.5 [g]	21.2 [g]	204 704 [d]	48.6 [g]	23.6 [g]
2002	443 821 [d]	68.9 [d]	98 082 [d]	42.8 [g]	23.5 [g]	200 007 [d]	52.1 [g]	22.3 [g]
2003	440 267 [d]	73.3 [d]	86 660 [d]	51.0 [g]	21.4 [g]	199 192 [d]	53.2 [g]	20.6 [g]
2004	467 985 [d]	70.7 [d]	96 081 [d]	44.2 [g]	24.2 [g]	204 073 [d]	51.6 [g]	21.4 [g]
2005	508 363 [d]	68.3 [d]	111 619 [d]	44.3 [g]	26.1 [g]	219 878 [d]	44.6 [g]	22.3 [g]
2006	528 525 [d]	66.6 [d]	120 259 [d]	42.2 [g]	27.1 [g]	228 898 [d]	42.8 [g]	24.6 [g]
2007	564 775 [d]	65.2 [d]	121 135 [d]	45.0 [g]	27.0 [g]	249 487 [d]	43.5 [g]	21.9 [g]

Note:

a Candidates who failed on the HG but who received SG passes.

b Candidates who failed on the SG but who received Lower Grade (LG) passes. Lower Grade Biology examinations were not offered.

c From Mukwevho, Khosa, and Kgobe (2004).

d From DoE (2002d, 2003b, 2004f, 200[5]a, 2005d, 2006b, 2009a).

e Figures provided by Umalusi (P. Mokilane [Statistician: SIR, Umalusi], personal communication, April 9, 2010).

f Figures were calculated from data supplied by Umalusi (P. Mokilane [Statistician: SIR, Umalusi], personal communication, April 9, 2010).

g Figures were re- calculated from data from DoE (2002d, 2003b, 2004f, 200[5]a, 2005d, 2006b, 2009a) because these sources included HG passes converted to SG passes in the HG pass rate and SG passes converted to LG passes in the SG pass rate. These conversions inflated the pass rates of candidates who wrote the examinations on each of HG and SG Grade.

Table 2.10 Number of students who wrote the IEB SC Biology HG and SG examinations from 1994 to 2007. Pass rates determined using adjusted marks not raw scores.

	Biology HG			Biology SG		
Year	**Number of candidates**	**Pass HG (%)**	**Pass SG[a] (%)**	**Number of candidates**	**Pass SG (%)**	**Pass LG[b] (%)**
1994	695 [c]	82.3 [d]	10.8 [d]	233 [c]	84.1 [d]	7.3 [d]
1995	774 [c]	81.9 [d]	11.6 [d]	280 [c]	80.4 [d]	11.1[d]
1996	1 528 [e]	87.2 [f]	10.7 [f]	513 [e]	88.3 [f]	8.2 [f]
1997	2 006 [e]	90.0 [f]	8.2 [f]	704 [e]	87.9 [f]	8.5 [f]
1998	2 246 [e]	91.1 [f]	8.3 [f]	623 [e]	93.3 [f]	5.1 [f]
1999	2 544 [e]	94.0 [f]	5.7 [f]	835 [e]	96.3 [f]	2.5 [f]
2000	2 811 [c]	91.5 [d]	8.0 [d]	804 [e]	94.3 [d]	4.6 [d]
2001	2 518 [e]	97.1 [f]	2.5 [f]	636 [e]	99.5 [f]	0.0 [f]
2002	2 671 [e]	94.5 [f]	3.7 [f]	677 [e]	99.4 [f]	0.0 [f]
2003	2 888 [e]	97.1 [f]	2.2 [f]	645 [e]	98.9 [f]	0.0 [f]
2004	2 981 [e]	98.5 [f]	1.2 [f]	643 [e]	98.9 [f]	0.0 [f]
2005	3 051 [e]	97.0 [f]	1.7 [f]	688 [e]	98.0 [f]	0.0 [f]
2006	3 164 [e]	98.6 [f]	1.0 [f]	714 [e]	99.4 [f]	0.0 [f]
2007	3 320 [e]	98.9 [f]	1.1 [f]	740 [e]	99.9 [f]	0.0 [f]

Note:

a Candidates who failed on the HG but who received SG passes.
b Candidates who failed on the SG but who received LG passes.
c Figures supplied by Umalusi (P. Mokilane [Statistician: SIR, Umalusi], personal communication, April 14, 2010).
d Percentages calculated from data provided by Umalusi (P. Mokilane [Statistician: SIR, Umalusi], personal communication, April 14, 2010).
e Figures supplied by IEB (E. Nel [Head of Assessment, IEB], personal communication, August 11, 2009).
f Percentages calculated from data provided by the IEB (E. Nel [Head of Assessment, IEB], personal communication, August 11, 2009).

of skills from those measured by languages, Mathematics and Physical Science curricula (Umalusi, 2004). Therefore Biology assessments, like the SC Biology examinations, can offer particular kinds of information as to what students know or understand or can communicate. Britton et al. (1996a) highlighted the general need for science examinations of world-wide college-bound students to reflect the importance of effective communication.

Post-1994, the SC performance of students from schools which were previously designated as disadvantaged[59] during the apartheid era has continued to lag behind that of their advantaged school counterparts (Orkin, 2006; Simkins, 2010) despite the fact that greatest gains were made in disadvantaged schools serving poor and working class children (Fleisch, 2008). A higher

[59] Schools administered by the DET, the HOD, the HOR and the homeland schools.

proportion of candidates from disadvantaged schools did not have as a home language English [or Afrikaans], the languages in which the SC examinations is set and written (DoE, 1998; Lolwana, 2006; Organisation for Economic Co-operation and Development [OECD], 2008; Umalusi, 2004). In addition, many students in advantaged schools received subject tuition in English or Afrikaans from Grade 1 (OECD, 2008). As a science, the language of Biology has a larger verbal component than mathematics and a smaller verbal component than a language subject (Abedi, 2002). This set of contrasts means that Biology offers a unique opportunity to start to explore the possible role of language in the performance of students in Biology SC examinations.

There have been suggestions that in the South African context, for students who take Biology at Grade 12 level, the mark obtained in the SC examination may be a good indicator of success in higher education study (T. Isaac, W. Fraser, personal communication, December 6, 2006; S. Muller, personal communication, July 2, 2010). If this inference is correct, what is it that Biology assesses, that might not be present in other subjects? Higher education stresses the development of higher intellectual skills, so one can speculate that possibly Biology assessments are more of a cognitive challenge than those of other subjects.[60] However, this interpretation is questionable given that Isaac (1990), in his study of the influence of the SC examinations on teaching in Indian Schools in Natal, found that the teaching of Biology concentrated mainly on the acquisition and recall of knowledge because the HOD SC Biology examinations stressed this level of activity. Similarly, Charoux (1993) found that the Transvaal Education Department SC Biology examinations were focussed mainly on the recall of knowledge. So what do Biology SC examinations question papers really measure? This study begins to explore this issue over an extended period of 14 years.

2.2.2 Why focus on Senior Certificate?

In South Africa, the SC[61] examination − commonly referred to as Matric − is written at the end of Grade 12 and signifies the end of secondary schooling. The SC comprises the only set of school examinations taken during twelve years of schooling that is both set and administered externally (i.e., the examination question papers are set and the answer scripts are marked by the examining bodies and not by the schools.) Quality assurance of these examinations was initially the responsibility of the JMB (1918 – 1992), then of Safcert (1992 – 2001), and now of Umalusi (2002 – present) (Lolwana, 2006). Therefore, the SC is the level of schooling that offers the same

60 Oberholzer (1995, p. 144) found that "[a]mong [H]igher [G]rade subjects [at Senior Certificate level], Physical Science and Mathematics and to a lesser extent Biology can be identified as 'severe' ". She makes no reference to SG Biology as either lenient or severe.

61 Called the NSC from 2008 onwards.

(or similar), moderated simultaneous assessment of all its candidates under controlled and monitored conditions. Because the SC examinations were public examinations, it was expected[62] that there would be good records of examinations scores over time for comparisons to be made of student achievement for different examination question papers.

The SC examination is one of certification that qualifies successful candidates either to continue with tertiary education or to enter the workforce (NEPI, 1992a). Therefore, it is important to have an indication of what knowledge and level of cognitive skills Grade 12 students are required to have (as communicated by the examinations) and hence exhibit when they leave the school system.

The South African 1999 SC Biology examination was claimed to be of an acceptable standard when it was benchmarked against comparable Scottish high-school-leaving examinations (DoE, 2000c). However, Crowe (2000) questioned these results because of the methodology which was used to perform the benchmarking exercise. Incorrect versions of the CBS document were used, and it was not clear which particular examining bodies examinations were submitted for benchmarking (Crowe, 2000). No empirical evidence, or methodology to generate the empirical evidence necessary, by which to compare the examinations of different examining bodies was provided.

2.2.3 Why select the period 1994 to 2007?

This timeframe covers a period of socio-political changes in the move to democracy in South Africa. The period beginning in 1994 was a period of intended transformation in all spheres of South African life, including the education system. The goals of transformed education in the post-1994 South Africa as stated in the Reconstruction and Development Programme policy framework (ANC, 1994), the White Paper on Education and Training (DoE, 1995) and the Tirisano Plan of Action (DoE, 1999, 2000b), were that quality education – especially in Mathematics and Science – be made available to all South Africans. For South Africa, the adoption of the principle of

[62] The author spent much time trying to track historical raw score data on the performance figures for Biology in the time period covered by this study. A Ministerial Committee Report (DoE, 1998) indicated that SC examination statistics were at that time managed by a company called QData. The author was unable to trace, via the Examinations Directorate of the national DoE, Umalusi, Dr Calitz (a previous CEO of Safcert), and current members of the Umalusi statistical team, the data concerning the raw scores of examining bodies or QData. Raw scores were never in the public domain for the period of this study. For this reason adjusted scores, combined for all state candidates have been included in this thesis, with an exception when reporting about the candidates' answer scripts where the author had direct access to the sample candidates' raw scores.

'education for all' has meant a major restructuring of the unequal education system inherited from the pre-1994 (apartheid) government.

Changes to the education system, relevant to this study included, for example, new examining bodies which replaced old ones and existing examining bodies which merged (Table 2.4), the introduction of new policies and interim syllabuses (Table 2.8), and changes which were made to the format of examinations with the introduction of a formative assessment mark for all students was implemented in 2001[63] (Table2.8). Linn (2001), based on research in the USA, indicated that typical test score patterns showed a decrease in student performance when a change was made to the design and structure of the assessment used. Linn found that student performance thereafter continued to increase until the format of the assessment changed again. The South African national SC pass rates decreased from 54% in 1996 to 47% in 1997 and thereafter increased to 73% in 2003 (Naidoo, 2006; wa Kivilu, 2006). The Biology results in government schools in 2001 when one examination question paper was split into two, did not show a decrease as might have been expected from Linn's (2001) work. Instead, student performance continued to rise. Why did that apparent improvement occur, and is the explanation in any way important to how we understand the standards of the SC Biology examinations?

Nuttall (1986, p. 126) described a study conducted in 1980 by Christie and Forrest which compared A-Level Chemistry examinations as "one of the most sophisticated British studies of examinations over time". The study was essentially declared invalid because the domain of chemistry was considered to have changed too much over the period of the study 1963 to 1973 so therefore no direct comparison could be made of the standards of the examinations (Nuttall, 1986). Choppin (1981, p. 5) revisited the Christie and Forrest study and found that "a mere 30 scripts from [the] 1963 [Chemistry examination] were re-marked according to 1973 standards" to test the viability of a methodology rather that to draw conclusions about changing standards. In Chapter 6, the author of this thesis will argue from the methodology developed and the resulting analyses conducted for this thesis that analyses of examination question paper should ideally take place over a *continuous number of years* to detect small incremental changes that might be apparent in both the policies and the examination question papers of those years, if inferences are to be drawn about changing standards. In addition, Dowd (2003) recommended that the content of assessments, especially that of public examinations, should be reviewed constantly and modified as and when needed – so changes over time might be expected. To select any two individual years separated by a period of time, such as 10 years in the British study previously described (Choppin, 1981; Nuttall, 1979),

63 Prior to 2001, some examining bodies included a year mark in the calculation of students' final marks but complete records could not be obtained of this practice.

means that small, or indeed potential incremental, changes in standards will not necessarily be discernable.

This thesis investigates the SC Biology examinations of fourteen continuous years which can be compared with one another. The first year of this study, 1994, was the first year for which the author could obtain a full set of HG Biology examination question papers.[64] Full sets of SG question papers were only available from 2001. Full sets of question papers per year were considered necessary to obtain an accurate reflection of the changes that occurred within and between years. A cut-off point of 2007 was necessitated because it was the last year that the 'old' Biology SC examinations were written.[65]

Different policies and combinations of the different interpretations of the policies (Table 2.6; Table 2.8) directed the examining of SC Biology at different times within the 1994 to 2007 period embraced by this study. These conditions allowed three continuous time periods: 1994 and 1995; 1996 to 2000; and 2001 to 2007, to be defensibly separated for comparative purposes. The period 1994 to 2007 therefore represents a particularly interesting period in the education system of South Africa. It allows for natural and meaningful comparisons between Grade 12 Biology examinations set, written and administered under different circumstances and different policies. These comparisons allow a better understanding of what characterized end-of-year Grade 12 Biology examination question papers and what these examinations might have meant in terms of student learning. This characterization can be used as a benchmark for comparing the new Life Sciences (which replaced Biology) NSC Grade 12 examinations, written for the first time in 2008. The conceptual framework and methodology developed in this study to explicate standards from SC Biology examinations provides an instrument to make comparisons between the SC Biology and the NSC Life Sciences examinations.

64 1994 also happens to mark the first year of a change in government in South Africa. Therefore 1994 was the start of a period of transition in all aspects of South African life, including education which adds to the richness of this study.

65 The context of the 2008 and 2009 examinations changed too much with the introduction of a new curriculum, for them to be included in this thesis. They are reported elsewhere (Crowe, in preparation).

2.2.4 Why use Western Cape Education Department candidates' answer scripts?

There was no archival system for Biology SC candidates' answer scripts beyond approximately six months after the corresponding examinations were written.[66] The author is resident in the Western Cape so it was easier to obtain the permission necessary to research the relevant WCED candidate answer scripts in 2005 and 2006, and the logistics of collecting and safe-keeping the scripts could be secured and guaranteed.

It might be argued that candidates' answers scripts from the WCED are not representative of what candidates' answer scripts would have been from the other state examining bodies because the WCED schools are better resourced and their students are then able to perform better. This argument might be valid if one considers that the WCED has the lowest proportion of non-fee-paying schools,[67] its schools had better access to sanitation and had more library space and its children had more access to schools than in any of the other provinces (Table 2.1). In addition the WCED schools had smaller classes than all provinces except for the Northern Cape and the Free State (Table 2.1). In terms of student achievement the WCED had more students that any of the other provinces who achieved literacy in the language of teaching and learning at the Grade 6 level, and more students who had completed Grade 9 than any provincial education department other than Gauteng, (Table 2.1). Also in support of the view that WCED is better resourced than the education systems of the other provinces is that, while the WCED did not have the greatest number of SC candidates in the two years for which this study examined candidates' answer scripts (i.e., 2005 and 2006), it had both the best overall SC pass rates both with and without ME (Table 2.11). Given this information, the results of the analyses of student performance data generated in this study should not be used to extrapolate to what was is happening in the whole of South Africa.

However, the WCED scripts do provide sources of information from the entire spectrum of performance from lower-performing students to higher-performing students, a range which might not have been available from some of the other provinces' education departments. As this study is a first in terms of a detailed analysis of candidates' SC Biology answer scripts, it was essential to analyze scripts which covered the entire range of abilities as reflected by the examinations.

66 The IEB have started archiving candidates' answer scripts as from the 2009 NSC examinations.

67 Van der Berg and Burger (2003) and Luckay (2010) used the amount of school fees as a measure of the socio-economic status of schools and found that it correlated with aspects of students' performance in Western Cape schools.

Table 2.11 A summary of SC examination results by province (government schools only), 2005 and 2006. Compiled from South African Institute of Race Relations (2008).

Province	Year	Number of candidates	Proportion who passed (%)	Proportion who passed with ME [a] (%)
Western Cape	2005	38 586	84	27
	2006	39 824	84	27
Northern Cape	2005	7 825	79	15
	2006	7 495	77	16
Gauteng	2005	76 202	75	21
	2006	73 216	78	23
Free State	2005	26 180	78	22
	2006	29 884	72	20
Kwazulu-Natal	2005	120 397	71	17
	2006	125 777	66	15
North West	2005	37 682	63	12
	2006	37 953	67	15
Mpumalanga	2005	38 811	59	13
	2006	39 040	65	14
Eastern Cape	2005	69 869	57	9
	2006	69 561	59	10
Limpopo [b]	2005	92 811	65	18
	2006	105 775	56	13
National	2005	508 363	68	17
	2006	528 525	67	16

Note:
a Matriculation endorsement.
b Formerly known as Northern Province.

2.3 Comparative international studies of science assessment systems including examinations

Several international and national studies have been carried out to understand assessment systems and the functioning of examinations as assessment tools. The studies are described here for two reasons. First, they each offer insight into what is methodologically important in the comparative analyses of examinations such as the SC examinations which are the focus of this thesis. Second, none of these studies used standards or validity as lenses to make comparisons of assessment questions nor did they analyze candidates' answer scripts, which necessitated the development of conceptual framework to underpin this comparative study of SC Biology examinations (Chapter 3).

In an attempt to contextualize the functioning of science assessment systems in the USA, Black and Wiliam (2004) studied the assessment systems, including the policies and practices of seven different countries - Australia (Queensland), France, Germany, Japan, New Zealand, Sweden and England. Policies with regard to assessment practices, such as internal assessment versus external examinations, were found to vary even between the provinces in one country, (e.g., in Canada, between Ontario and Quebec, and in Australia between Queensland and New South Wales) (Strachan, 2002). A general lack of coherence between different assessments administered within a single education system has been attributed, in part, to policies and practices that do not recognize that each assessment and its purpose need to be closely linked (Rutherford, 2003). In a coherent assessment system, different assessment strategies are designed to address different questions and produce results which complement one another (National Research Council [NRC], 2006) and the of content of assessments, especially that of public examinations, should be reviewed constantly and modified as and when needed (Dowd, 2003). Black and Wiliam (2004) and Noah (1996) concluded that there was no one 'right' assessment system that would satisfy all the different countries' needs. Black and Wiliam (2004) proposed and discussed several models of assessment, including examinations, that incorporated what they believed were the best features of each of the practices they studied.

Britton and Raizen (1996) presented a collection of scholarly comparisons of the end-of-secondary-school examinations in mathematics and science for college-bound students in England and Wales, France, Germany, Israel, Japan, Sweden and the United States. Eckstein and Noah's (1993) perspectives on the policies and practices of secondary school examinations contextualized Britton's and Raizen's work. The Britton and Raizen study was intended to offer USA policymakers and educational organizations a better understanding of what constitutes curricula of an international calibre through the lens of secondary school examinations. Because their study was international, they adopted and augmented a version of the TIMSS framework (Robitaille, McKnight, Schmidt, Britton, Raizen & Nicol, 1993) which was in use at the time for their comparisons.

The TIMSS framework was adopted in two other USA studies of international curriculum practices. To explain cross-national differences in educational achievement progress, Valverde (2000) described patterns of underlying strategic themes in curriculum policy exhibited by various countries. Valverde and Schmidt (2000) extended this study by comparing the policy and curricula of countries that outperform the USA in the international TIMSS tests, in an attempt to understand what is meant by 'world-class standards' in mathematics and science.

Current educational practices, including the assessment practices, of some countries in Africa have been compared to the corresponding educational practices of the countries which historically influenced each of their education systems, in an attempt to understand the quality of the education they offer. Lewin and Dunne (2000) observed convergence between the assessment practice, particularly in primary school science public examinations, of nine Anglophone countries. They attributed this convergence to the structural similarities between the African education systems studied, rather than to the influence of international assessment reform initiatives and practices of, say, England, especially. Elsewhere in Africa, Valverde (2005) considered the mathematics and biology examinations which signalled, in part, the exit from secondary schooling in some Middle East and North African (MENA) countries. Valverde developed "signatures" of the different examinations and used these signatures to compare the content of the examinations in each country to the secondary school exit examinations in France and to discuss each country's results of the National Assessment of Educational Progress (NAEP) test, in an attempt to inform curriculum development and educational reform in the MENA countries that he studied. However, Valverde made no attempt to link his findings to student achievement in any of the examinations that he studied, nor did he describe these examinations he studied in terms of 'standards' or learning outcomes or in terms of cognitive demand. The MENA countries are particularly interesting for this study because they face similar challenges to South Africa, in that most of the region is characterized by "the dearth of good data to guide educational reform" (Valverde, 2005, p. 31) and to understand school-leaving examination systems

In an attempt to understand the quality of the South African SC, Umalusi (2007) compared the syllabus and secondary school exit examinations of four African countries, that is, Ghana, Kenya, Zambia and South Africa for some subjects, including Biology. The analysis of examination question papers compared the structure of the 2004 examination papers from each of the four countries and their cognitive demand as defined by the Revised Bloom's Taxonomy (Anderson , Kwathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths, & Wittrock, 2001). This intra-African study found that South African examination papers tested a wider range of cognitive skills and knowledge dimensions than the other countries, but the analysis covered only one year of examinations from each country. In addition, the South African examinations were found to include more questions which require reading and visual interpretation that the examinations of the other countries studied (Umalusi, 2007).

The studies described above compared science assessment systems within and between different countries. While some of these studies made reference to standards, none made use standards as a lens to understand the examinations or make comparisons between different examinations; neither were student answer scripts analyzed in their research. However, these studies highlighted the kinds of variables which are important in analyses of examination question papers. Consequently, methodologies from some of these studies have been incorporated into the operationalization of the conceptual framework (Chapters 4 and 5) developed for this study in Chapter 3, to the generate the standards of SC examinations in Chapter 6.

2.4 **Chapter summary**

This chapter described the context in which the SC Biology examinations took place between 1994 and 2007, a period of transformation in South African education. The introduction to this chapter summarizes the inequities that exist in South African schools. While the social milieu in which the SC examinations operate in South Africa is not a focus of this thesis, this introduction serves to orientate the reader to the complexities of the South African context. The next section wove together information from a number of disparate sources, both from the public and private domains, to describe in general the processes, events and procedures associated with the examination of SC Biology examinations. This discussion was followed by an explanation for both the areas chosen to be the research focus of this study and the delimitations of this study. The selection for study of: Biology; examinations at the SC level; the timeframe of 1994 to 2007; and WCED candidate answer scripts, are supported by evidence. The supporting evidence comprises: a compilation of SC pass rates; SC Biology pass rates; documentation of the polices which governed the SC Biology examinations during this period of time. Because the SC Biology examined the work of only one year, the final year of high school the CBS which covers the last three years of high school is positioned within a international TIMMS high-school leaving Biology syllabus framework. The international comparative studies of science school leaving examinations described provided insight into those variables that should be considered when comparing examinations. None of these studies used standards as a lens to view the examinations which necessitated the development of a conceptual framework to understand the standards implicit in the SC Biology examinations in the following chapter.

CHAPTER 3

DEVELOPMENT OF A CONCEPTUAL FRAMEWORK FOR DESCRIBING AND ANALYZING STANDARDS OF SOUTH AFRICAN SENIOR CERTIFICATE BIOLOGY EXAMINATIONS

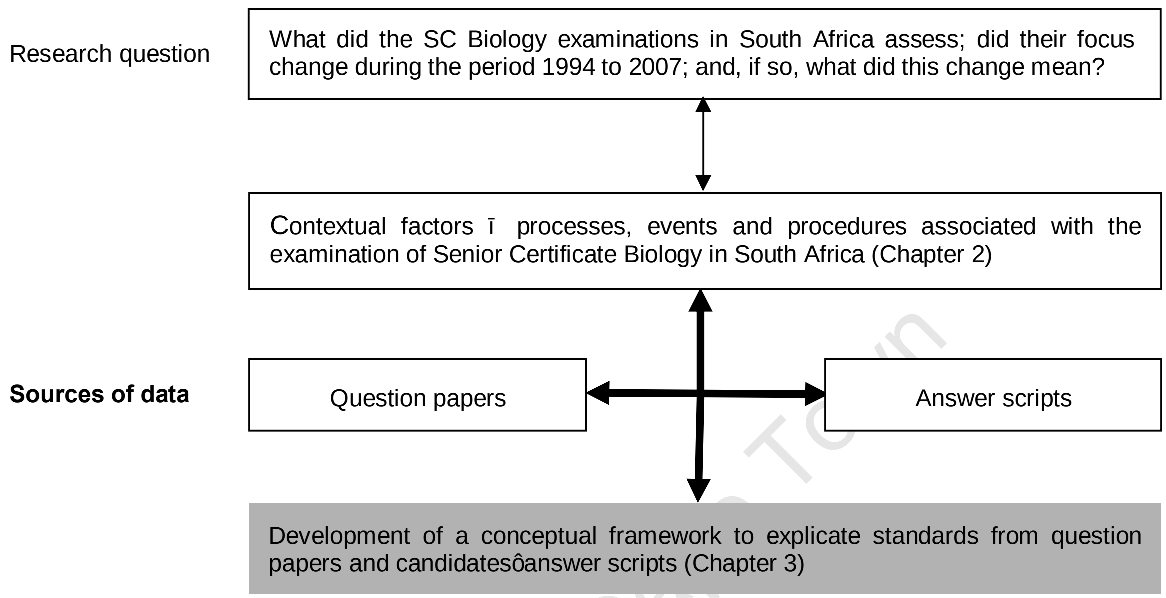


Figure 3.1 The relationship between Chapter 3 and the previous chapters.

Potentially assessments, like examinations, "are powerful instruments for educational change" (Shavelson & Huang, 2003, p. 11), provided their results are interpreted in a meaningful way. South Africa does not offer a standards-based curriculum, and so what competencies the SC examinations measure or do not measure, in terms of standards, is unclear. Without information about competencies, we cannot interpret students' achievement in the examinations (Nitko & Brookhart, 2007), nor can we begin to understand what changes in student achievement between years might mean as the "cost of not having national standards is the cost of non-comparability" (Noah, 1989, p. 18).

Umalusi (2004)─the quality assuror of standards in the South African education system─and the *Marking Matric* project (Reddy, 2006a) have indicated the need for a better understanding of SC examinations results and more specifically, what they can reveal about the state of education in South Africa. If "a nation's [South Africa's] educational standards are embodied in its secondary-school-leaving examinations" (Eckstein & Noah 1993, p. 143), we need to begin to understand what information about standards our SC examinations convey to society. Only then can we begin to understand what our expectations of schools are, whether these expectations are reasonable, which expectations need re-thinking, and what the implications of an understanding of standards are for policy. So where do we begin to look for the meaning of the SC examinations in terms of standards and student competencies?

Mislevy, Steinberg and Almond (2003, p. 8) posited that "[v]iewing assessment from the perspective of arguments is natural once one accepts that validity, [is] the cardinal virtue of assessment" and is about "the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment" (Messick, 1989a, p. 13, emphases in original). Assessment data then acquires meaning in relation to the particular inference being made. Simply put, tests scores acquire there meaning through validity (Cizek, Rosenberg & Koons, 2008).

Umalusi certified as successful those students who wrote SC Biology examinations and who achieved particular symbols in the examinations (Chapter 2). Implicit in this certification must have been the belief that the standards of the SC examinations were acceptable. In Messick's (1989a) terms, what empirical evidence was there to support, or not support, the certification of the SC candidates as having achieved a particular symbol on the HG or on the SG, or meeting the required standard? There is little empirical evidence about the meaning of student performance in any of the subjects that were offered in the SC examinations, especially with regard to standards (Chapters 1 and 2). No studies which view high school leaving science examinations through the lens of standards could be found in the literature.

The absence of well-articulated, and commonly understood, educational standards with respect to the SC examinations necessitated an explicit explanation of, or argument for, how standards in the context of SC Biology examinations could be understood, through the use of validity evidence, and analyses of SC examinations. Understanding the SC Biology examinations through the dual lenses of validity evidence and of standards required that the meaning of, or use of, the concepts 'standards' and 'validity' be explored. This chapter provides a conceptual framework, incorporating validity evidence and standards, to guide the unpacking of the standards inherent in SC Biology examinations.

The first part of this chapter develops a conceptual understanding of the international use of standards in the context of education. The intention here is not to uncritically assume that use of educational standards elsewhere in the world can be, or should be, applied 'as is' to a South African context. The role of assessments in standards-based education systems is discussed through a particular broad conceptualization of curriculum because of the coherence that standards potentially bring to education systems, provided they are explicated clearly. If explicated before examinations are set, the task of determining if an examination has met the standards is easier to do. Validity brings meaning to student performance in an assessment because the validation of the assessment provides evidence to support inferences and decisions, that are made about students who took the assessment. Validity and types of validity evidence, including alignment methodologies and the equating of assessments is discussed thereafter.

What follows is a description of a formal approach to understanding─a conceptual framework─how standards of the SC Biology examinations might be explicated, described and analyzed (Chapter 5 and 6) in ways that are appropriate to unique South African educational contexts (Chapter 2), in order to answer the specific research question and sub-questions posed in this thesis (Chapter 1) . The way in which the conceptual framework developed here is to be used in this study is consistent with two functions for theoretical frameworks given by Wagner (2003, p. 88-89): "The first is to provide organization to[,] and a description of [,] the existing landscape. The second is the heuristic function of enabling us to move beyond the boundaries of the existing landscape".

This conceptual framework 'built' in this chapter links examinations and standards using the relationships between standards, validity and assessments, and the role of standards in curriculum as explored earlier in this chapter. The significance of the conceptual framework is described in this chapter, rather than in the last chapter of the thesis, because this conceptual framework guides the remainder of this thesis. That is, the framework proposes how the standards implicit in the SC Biology examinations are conceptualized, operationalized, explicated and compared from analyzes of the SC Biology question papers and answer scripts. This chapter concludes with working definitions of content standards and performance standards that are used throughout the remainder of this thesis.

There is unavoidable overlap between some of the sections of this chapter. For example, there is overlap between the section which conceptualizes educational standards and the section on standards in the USA, because most of the explanations of what standards are come from USA literature. Similarly, between the section on standards in the USA and the section explaining what it means to have a standards-based curriculum, while one section serves to describe the USA practices of standards and explain how they came into being, the other section serves to clarify the function of standards in a curriculum. At first glance the section on validity may seem unnecessarily lengthy, but it is an understanding of the role of validity in bringing meaning to test scores that justifies the decision to use validity evidence to generate the implicit standards of the SC Biology examinations in this study.

3.1 Standards in education

3.1.1 What are educational standards?[68]

A review of the research literature on assessment indicates difficulties, misunderstanding and confusion in how terms describing educational measurement concepts, and the relationships between them, are used (Frisbie, 2005). One of these terms – 'standards' – has multiple uses in different components of education systems, such as, for example, in assessment, curriculum, teaching and textbooks and in written discussions (Linn & Baker, 1995). Whether a single definition of educational 'standards' is possible or desirable, especially with respect to examinations, has been challenged elsewhere (Baird, Cresswell & Newton, 2000) since "[s]tandards are not anyone's truths" (Brennan, 2001, p.13). The absence of an explicitly stated standards-based curriculum in South Africa made it essential to develop a general understanding of the meaning of the concept 'standards' as it applies globally to educational issues as the starting point for this chapter. In this section, an explicit meaning is constructed for standards. What follows is a review and synthesis of educational standards as practiced in the USA and in the UK, the Netherlands and parts of Australia. These countries represent the only countries for which the author was able to find published literature which clearly documents how the particular country developed its own specific understanding and practice of educational standards. Moreover, the USA and collectively the UK, the Netherlands and Australia offer different conceptualizations of standards which, once they are understood, were modified for use in the context of South African SC examinations practice. In addition, a study of the development of a standards-based curriculum over time in the USA demonstrates the value that explicit standards can ideally bring to an education system as a whole.

In order to develop a working definition of 'standard(s)' to guide this study, descriptions of 'standard(s)' from a range of sources were sought. One of these sources was a well-known English dictionary (Fowler & Fowler, 1965) whose explanation of standards preceded its widespread use in education. Other sources, perhaps the most extensive descriptions of any educational standards at a secondary school level are the science standards found in two important USA science teaching reform documents, namely, the American Association for the Advancement of Science's (AAAS) *Benchmarks for Science Literacy* (AAAS, 1993) and the *National Science Education Standards* compiled by the NRC (NRC, 1996). Both the AAAS and the NRC are comprised of people who represent stakeholders at various levels of scientific enterprise and science education in the USA. These two documents embodied the standards that these two groups of people value as important in science education.

68 'Educational standards' are considered to be synonymous with 'academic standards' in this thesis.

Fowler and Fowler (1965, p. 812) defined ‘standard(s)’ as

> [a] **specimen** or **specification** by which the **qualities required** of something may be **tested**, required degree of some **quality**, **level reached** by average specimens, (attrib.) serving as **test**, correspond to the s.[tandard] of **recognized authority** or **prevalence**. [Emphasis added]

Benchmarks for Science Literacy (AAAS, 1993, p. 322) defined ‘standard(s)’ with respect to Science (including Biology) education as

> [a] standard, in its broadest sense, is **something** against which other things can be **compared** for the purpose of **determining accuracy**, **estimating quantity**, or **judging quality**. In **practice,** standards may take the form of **requirements** established by an **authority,** indicators such as **test scores**, or **operating norms** approved of and fostered by a **profession**. [Emphasis added]

The *National Science Education Standards*, a document of the NRC of the USA, includes the science standards negotiated by stakeholders as being important in science education, but does not specify who should construct the standard(s). The NRC definition (NRC, 1996, p. 12) expands the AAAS definition by specifying exactly where in the practice or profession of education the standard(s) should operate. Accordingly,

> [t]he term “standard” has multiple meanings. Science education standards are **criteria** to **judge quality**: the quality of what **students know and are able to do**; the quality of the science **programs** that provide the **opportunity for students to learn** science: the quality of science **teaching**; the **quality** of the **system** that **supports** science **teachers** and **programs**; and the **quality** of **assessment practices and policies**. Science education standards provide **criteria** to **judge progress toward** a national vision of learning and teaching science. [Emphasis added]

Collectively the definitions from the three sources highlight at least six aspects of standard(s) that can be recognized within the context of education, and indicated in bold below. Bracketed words are the exact words or phrases from Fowler and Fowler (1965), AAAS (1993), and NRC (1996).

Standard(s):

1. are determined by a **recognized authority** (recognized authority, authority, profession);
2. are specified **criteria** (specimen, specification, qualities required, something, requirements, criteria, operating norms, students know and are able to do);
3. should be **practiced** at all levels of an education system, that is, teaching, learning and assessment (prevalence, practice, system, supports, programs, teachers, opportunities for students to learn, assessment practices and policies);

4. can be **measured** (determining accuracy, estimating quantity, tested, test scores, operating norms, students know and are able to do); [69]

5. signify **achievement** (level reached, judge, judging, progress toward, compared, operating norms, students know and are able to do); and

6. denote **quality** (quality).[70]

Other definitions of educational standards have been formulated. For example, Cresswell (1998, p. 257) proposed a definition of standards within the contexts of examinations as "the value accorded to students' work by judges accepted by all interested parties as competent to make such judgments". Cresswell's definition encompassed aspects of standards related to **recognized authority**, student **achievement**, **quality** (value) and **measurement,** but made no explicit mention of **criteria** or of **practice** outside of the work of students. Yet standards do not function independently of criteria (Sadler, 1996) or of practice in teaching and in learning (Noah, 1989). Cresswell's work focused on how students' perform in a specific examination rather than on the features of the examination itself (Baird, Cresswell & Newton, 2000). However, Cresswell's definition relates to a context (i.e., the UK) where judgments about the standards of examinations are made by experts (Baird, 2000; Baird et al., 2000), which by implication means that the experts have a specialized understanding of both criteria and practice to inform their judgments. Broadfoot (1979, p. 12) described experts as "people with specific qualifications and experience, who are entrusted with making judgments which conform to prevailing social values about desirable achievements". Others, such as Wurtz and Malcom (1993) and Collins (1998) argued that both the identification and inclusion of all relevant stakeholders were vital to establishing or constructing standards. For example, in his discussion of the USA's *National Science Education Standards*, Collins (1998) included the country's political leadership and the public as stakeholders. Thus, while the development of standards should be informed by societal expectations and the visions of

[69] "[S]ocial scientists can measure anything that exists" (Babbie & Mouton, 2009, p. 108) provided that what is being measured is conceptualized, defined and operationalized. This chapter conceptualizes 'standards' and Chapters 4 and 5 operationalize 'standards' for this study. This author is cogniscent that the use of measure (and measurement) in this way is contested in the literature (e.g., Humphry, 2011; Michell, 1999, 2008).

[70] In South Africa, post-1994 education policies repeatedly referred to 'quality' without qualifying what is meant by the term. Jansen in his critique of the implementation of education policy writes of his efforts in seeking "theoretical explanations of quality" (Jansen, 2001, p. 271). Similarly, Porter, McMaken, Hwang and Yang (2011a, p. 186) write that "[q]uality is difficult to define and assess". The definition constructed for standards in this thesis uses 'denote' so that 'standards' and 'quality' be viewed as synonymous for this study. This view does not include any inherent questioning or judgement as to whether the standards are good, bad or world-class. However, this study specifies that standards are developed by the society in which they function. It also offers a cogent explanation for 'quality' with respect to the South African curriculum if viewed from the perspective of a standards-based curriculum (see Sections 3.1.4 and 3.1.5).

professionals in the field, past practices and research information should be part of the process (Bybee, 2009).

Cresswell (1998) argued further that acceptance of his definition of standards, and the notion that 'standards' is a social construct (i.e., a concept that is created collectively by humans [McMillan & Schumacher, 2001]), means that objective comparisons over time cannot be made and that "a truly objective educational assessment system cannot exist" (p. 258). This view contrasts with that of Noah (1989) who argued that national standards are necessary if comparisons are to be made of assessments within a country. However, the concept of a social construct can be explicated and examined, what becomes important are such aspects as: "Whose social construct?" and "What evidence formed that social construct"? Moreover, provided that the social construct can be explicated it can be used to make comparisons, as changes should be obvious through that lens.

Embracing the argument that standards are a social construct which should benefit society through "the bolstering of public trust and confidence in educational, medical, professional and vocational areas" (Cizek & Bunch, 2007, p. 8), the notion of at least some constancy in standards, and therefore comparability in standards, requires that the aspects of standards specified above be modified. Moreover, Bybee (2009) argued that in addition to societal expectations and the vision of experts in the field, past practices and research should inform the development and maintenance of standards. Accordingly, the first aspect of standards is reworded and a notion of regular revision is added at position two. Therefore,

Standard(s):

1. are determined by **recognized authority through consensus between stakeholders**
2. are **reviewed regularly and revised if necessary** as more is learned about how they function within an education system;[71]
3. are specified criteria;
4. should be practiced (attained) at all levels of the education system;
5. can be measured;
6. signify achievement; and
7. denote quality.

[71] Creating and reaching consensus on standards is both difficult and time-consuming (NRC, 2008) and it was recommended that revision should happen at least every 10 years to embrace the current societal viewpoint of what a student of science should know (NRC, 2006) − "[k]nowledge grows and so should we" (Wineburg, 1997, p.255).

It is helpful to separate these seven aspects of educational standards into two groups based on how they are recognized by society, namely

1. the group of characteristics of standards which concern the establishment and maintenance of the standards[72] (determined by recognized authority through consensus between stakeholders; reviewed regularly, revised if necessary); and
2. the group of characteristics concerned with how the standards are operationalized (specified criteria; practiced; measured; signify achievement; denote quality).

The inter-relationship of these two groups of aspects of educational standards, namely, those concerned with the establishment and maintenance of standards and how they are practiced, was highlighted in recent conversations about the newly developed Common Core State Standards in Mathematics and English Language Arts and Reading in the USA (Beach, 2011; Cobb & Jackson, 2011; Porter, McMaken, Hwang, & Yang, 2011a, 2011b). Clear, shared, and negotiated conceptualizations of what standards are, how they are operationalized[73] and how they can be measured, are crucial to understanding the role of standards in education systems and how they are actualized[74] in teaching and learning (Porter et al., 2011b).

There are seemingly simpler definitions of standards than the one constructed in this study. However, these definitions do not include all the aspects in the constructed definition. For example, standards are "statements about what students should know or be able to do [**criteria, achievement**] as a result of schooling [**practiced**]" and which define the performance [**achievement**] to be accepted as evidence that learning has taken place [**measured**] (Griffith, 2008, p.100). Nitko and Brookhart (2007, p. 522) have an even simpler definition: "Statements about what students are meant to know [**criteria, achievement**]". Both Griffith's and Nitko and Brookhart's definitions focus on a specific kind of educational standards recognized in science education, that is, *content standards,* which are a set of knowledge, skills (Herman & Webb, 2007) or abilities (Cizek & Bunch, 2007), that students are expected to have at particular times within the education system or, more simply, what it is that students should be expected to know or do

72 The author deliberately does not use the term 'standard setting' because this term is used in the literature in two different ways (see Section 3.1.2).
73 'Operationalized' refers to how standards as conceptualized here are actualized for measurement.
74 'Actualized' is used here to mean how standards are practiced and contextualized.

(Schmidt, Wang & McKnight, 2005), or to have mastered (Cizek & Bunch, 2007).[75,76] Wurtz and Malcom (1993) and the NRC (2006) stated that content standards should also clearly and explicitly describe what it is that students need to do with the required knowledge and skills or the performance expectations.[77] The content standards define the domain over which teaching and assessment take place (Schafer, 2011), and defining the domain is critical to the measurement of change (Nuttall, 1986). *Performance standards*[78] describe how students demonstrate their competence, proficiency or mastery in the subject matter defined by specific content standards (Herman & Webb, 2007; NRC, 2006) and explicate what information will be accepted as evidence (NRC, 2006) to define levels of achievement considered acceptable or outstanding (Linn & Baker, 1995, NRC 2006). That is, performance standards define the nature and quality of student performance required to determine "how good is good enough" and whether the content standards have been met or not (Wurtz & Malcom, 1993, p. iii).

"Performance standards are arguably one of the most controversial topics in educational measurement" (Linn, 2003, p.1). Performance standards appear in different forms in the research literature: namely, either as statements of what students should be able to do at particular levels of competency (Hansche, 1998; Nitko & Brookhart, 2007) together with the nature of the evidence required to demonstrate that competency (Linn & Baker, 1995), or as a cut-score[79] that separates achievers at different levels or competencies (NRC, 2006).[80] The routine use of performance standard as a synonym for cut-score is considered ''less-than-accurate" (Cizek & Bunch, 2007, p. 16). Irrespective of the conceptualization of performance standards used in particular contexts, definitions of content standards and performance standards remain distinct but interrelated, with performance standards being derived from established content standards (Linn & Baker, 1995). However, content standards, and especially performance standards, have been operationalized in subtly different ways in different countries and regions of the world, as is discussed later in this chapter.

Linn and Baker (1995) recognized a third kind of educational standard called *opportunity-to-learn* (OTL) *standards* which defined criteria for assessing adequacy of the learning opportunities of

75 Bloom, Hastings and Madaus (1971) described what are now called 'content standards' as 'educational objectives' which included two components: a content (subject matter) component and a behavioural component (what a student was required to do with the content).
76 Content standards were born out of curriculum standards (McClure, 2005).
77 Schmidt and McKnight (1995, p.341) defined *performance expectations* as "what students are expected to do with the [subject-matter] topic" and *topic* as content.
78 Also known as achievement standards (NRC, 2006).
79 Also spelled 'cutscore' and 'cut score' in the literature.
80 Some earlier worked considered what students could do with knowledge, i.e. the skills, as performance standards (Wurtz & Malcom, 1993).

students. Porter (1995, p. 21) believed that OTL standards “have become one of the most contentious issues to come on the education scene in some time”. While OTL standards are not a focus of this study they will be briefly discussed in Section 3.1.5, because of the role their role in a standards-based curriculum.

A review of the literature about educational standards revealed an array of sources exploring such questions as how standards are established, maintained and operationalized by different countries. Examination of standards-related practices in education provided insight into ways in which some aspects of these practices could be applicable to the South African context. Resnick, Nolan and Resnick (1995, 1996) and Louis and Versloot (1996) demonstrated some of the problems associated with comparative studies of the role of standards in different education systems, and stressed the importance of understanding the links between practice and context in education systems, when drawing comparisons about standards in different countries. Articulating how standards are understood, and how they have influenced practice differently in the USA and in the UK, offers two different perspectives of “a rich treasury of educational experiences on [either] side of the Atlantic” (Bruce, 1991, p. 33). Studies of educational standards and the contexts of these countries therefore offer insights into ways in which aspects of this current study in South Africa ─ a country which does not have formally stated standards ─ might be used to constitute or infer standards. For this reason, and because most of the research literature about standards comes from the research communities in those countries, the two contexts are reviewed separately below. Other studies from the Netherlands and parts of Australia are included with the UK, where appropriate, because of the similarity in the modes by which these countries practice performance standards.

3.1.2 Standards in the United States of America

The contemporary USA science education system is characterized by the centrality of standards and their influence on the education system as a whole (NRC, 2006). The development of mathematics and science standards in the USA was orchestrated in response to societal needs to improve the quality of teaching and learning of science in that country (Bybee & Ferrini-Mundy, 1997). The standards are used as a reference point against which both qualitative and quantitative evaluations of the various components of education systems can be made (NRC, 2006; Porter, 2002). In order to understand the USA standards-based education system, it is necessary to examine the history of standards in this country.

When the paradigm, standards-based education, was emerging in the USA, Finn (1990, p. 586, emphases in original) contrasted the shift from educational "inputs" to educational "outputs" (or outcomes) in the following way:

> Under the *old* conception … education was thought of as process and system, effort and intention, investment and hope. To improve education meant to try harder, to engage in more activity, to magnify one's plans, to give people more services, and to become more efficient in delivering them.
>
> Under the *new* definition, now struggling to be born, education is the result achieved, the learning that takes root when the process has been effective. *Only* if the process succeeds and learning occurs will we say that *education* happened. Absent evidence of such a result, there is no education − however many attempts have been made, resources deployed, or energies expended.

Finn (1990) challenged the direct and causal relationship between inputs and outcomes that was assumed in the 'old' paradigm, citing the Coleman Report which was released in the USA in 1966. The Coleman Report concluded that in the USA, if educational equality was conceived of in terms of what students actually learned as opposed to the time, money and energy expended, the input variables might not have much to do with educational equality, if measured by student learning. Later, Fuhrman (2001, cited by Ferrini-Mundy, 2004, p. 26) described standards-based reforms as "reforms intended to anchor key aspects of policy─curriculum assessment, teacher education, and professional development─around policy level statements [standards] of what students should know and be able to do". Using a slightly different argument, that because in a standards-based system standards are the inputs *and* they define the output, Bybee (2009, p.6) considered national standards to have the "potential capacity [and therefore the power] to change the fundamental components of the [USA] education system at a scale that makes a difference". Such views of standards became intertwined with systemic reform which embraced the idea that the functioning of an educational system is dependent on the coherence or *alignment* (see Section 3.3) among the various parts of the education system (Smith & O'Day, 2001, cited by Ferrini-Mundy, 2004). Coherent, standards-based, education systems "do not develop by accident; they must be deliberately designed so that all the measures work together both conceptually and operationally" (NRC, 2006, p. 5). In a standards-based curriculum, standards drive the curriculum (NRC, 2006). More specifically, McClure (2005, p. 7) described *content standards* as curriculum standards because they were the "basic building blocks of a[n educational] systemic reform effort known as the standards movement" and *performance standards* as the standards which would "establish and describe the kind of student work that met, or fell short of the standard". McClure (2005) described the two objectives of this reform effort as equity (i.e., standards applied to all children and all schools) and to raise the overall performance of all students in terms of what they know and can do.[81] This view meant that standards for students were a starting point for what teachers had to know and be able to do, and a lever for the professional development of teachers (McClure, 2005; Porter & Smithson, 2001).[82] Such an education system would be described as having a standards-based curriculum where the standards act as an organizing principle or the framework which makes the requirements of education transparent to all stakeholders (McClure, 2005). In a standards-based system, the standards "are common benchmarks frequently incorporated into teachers' rationales" (Penick, & Harris, 2005, p. 7). In addition to being coherent, policies about standards should be clear and promote challenging content as part of the curriculum (Bybee, 2009) and, in the case of national standards, should be considered a resource that allows for regional autonomy and how they are used (Bybee & Ferrini-Mundy, 1997).[83]

Global awareness of, and research into, the importance of science (and mathematics) in education increased during the latter half of the last century, and the chronology and implementation strategy of the science standards in the USA has been well documented. The AAAS and the National Academy of Sciences (NAS) adopted a position on just how crucial science (and mathematics) education is to having functional, globally competitive citizens in the USA in the twenty-first century. This position was captured in two documents: *Science for all Americans* (AAAS, 1990) and *Benchmarks for Science Literacy* (AAAS, 1993). The NRC used these two documents to develop─in extensive consultation with interested stakeholders, such as professional scientists, education researchers, teachers and employers─a set of standards, that is, the *National Science Education Standards* (NRC, 1996). These standards detailed what students should learn and be taught, from Kindergarten through to Grade 12, in order to achieve the country's science education goals. The connections between these three major documents were well established (e.g., Raizen, 1997) and a strategic framework to implement standards-based reform in science education through these documents was designed (NRC, 1997). The framework developed by Bybee, Ferrini-Mundy, and Loucks-Horsley (1997) described five dimensions important in the implementation of standards and ensured a positive and effective influence of the standards (NRC, 2002). These dimensions represent the way in which the current science standards were introduced into the USA education system. The dimensions are: a) dissemination─developing an awareness of the science

81 However, in principle standards bring equal access to students but does not guarantee equality of outcomes (Robinson, 2011).

82 In his critique of the *National Science Education Standards* (NRC, 1996), Rodriguez (1997) argued that if the standards were to deliver on their promise to transform science teaching they should have provided conceptual guidance to users of the standards by articulating the theoretical frameworks and empirical evidence on which recommendations for change were based. In South Africa theoretical frameworks and empirical evidence supporting the selection of content, the emphases of content, and assessment decisions were conspicuously absent from the syllabus and guideline documents, the *de facto* standards, analyzed in this study, and post-2007 remain absent from similar NSC documents.

83 Wiggins (2011) challenged policymakers who determine standards for high schools in the USA to think outside of traditional notions of what is important for students to learn, to move away from "whose standards" and to rather ask "which standards" (p.30) will best serve high school students in their futures.

standards documents; b) interpretation─increasing understanding and support for standards; c) implementation─changing policies, programs and practices to be consistent with the standards; d) evaluation─monitoring the impact of new policies, programs and practices; and e) revision─improving the efficacy and influence of standards (NRC, 1997) (Figure 3.2). The Science Standards are currently being reviewed and revised in the USA (Robelen, 2010) and a new conceptual framework to guide these standards has recently been completed (NRC, 2012).

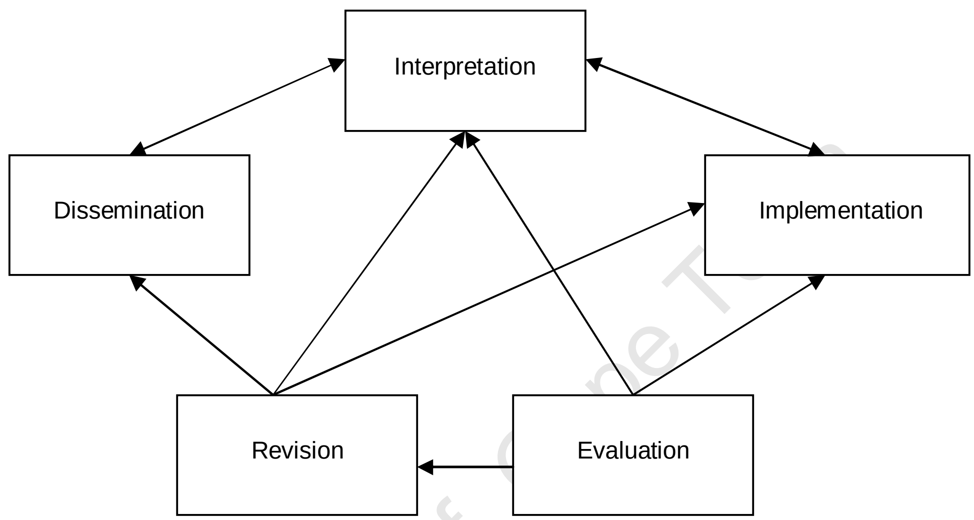


Figure 3.2 Relationships among the dimensions of the strategic framework for the implementation and the maintenance of standards in the USA (Bybee et al., 1997).

Another sense in which 'standards' are used in assessment in the USA is as achievement standards. Because achievement standards are always used with reference to student performance, they are also known as performance standards (Herman & Webb, 2007). However achievement/ performance standards are inconsistently used in conceptually different ways (NRC, 2006) and there is "no agreed upon best method" (Linn, 2003, Controversy Regarding Performance Standards, para. 2) for determining them. Test developers and psychometricians use performance standards to separate one level of assessment achievement from another (Cizek & Bunch, 2007). Cizek and Bunch (2007) defined a performance standard as "the abstract conceptualization of the minimum level of performance distinguishing examinees who possess an acceptable level of knowledge, skill or ability necessary to be assigned to a category, or for some other specific purpose, and those who do not possess that level" (p. 337) and use "the term performance standard cut-score essentially interchangeably with terms such as cut-score, standard, passing score, and so on" (p. 14). Those concerned with curriculum delivery such as, for example, teachers, use the term in relation to student proficiency at particular levels of performance (NRC, 2006). Kane (1994, p. 426) described a performance standard as the "conceptual version of the desired competence, and the passing score is the operational version". The development of measures of how well a student performs, that is, achievement/ performance standards for an assessment/examination, is called *standard[-]setting* (Cizek & Bunch, 2007). "[S]tandard setting is an integral part of most test use, no matter how formal or informal the assessment may be. We tend to reserve the term "standard-setting", however for formalized procedures associated with high stakes and/or high visibility testing programs" (Brennan, 2001, p. 13).[84]

Rogers and Dawber (2002) advocated that both perspectives of performance standards described above, that is, student proficiency at specific levels and criteria for how these different levels of performance are distinguished, be considered when standard-setting, alongside the relationship between performance standards and the content standards. In other words, performance standards communicate how well examinees are expected to perform in relation to the content standards (Linn & Herman, 1997). A performance standard then becomes a statement defining and describing the knowledge or skills that students must show at particular performance levels, which are differentiated from each other by cut-scores (Hansche, 1998). One function of standard-setting, within an examination, is to provide clear rules by which examinees can be assigned to one of several ordinal categories of performance (standards) based on their test score (Jaeger, 1991), that is setting the cut-scores. The cut-score is operationally defined as the score on the score scale which separates adjacent ordinal performance categories, and cut-scores are usually developed *after* the assessment is administered (NRC, 2006). Setting cut-scores are considered one of the most important processes in testing because of the consequences associated with the resulting classifications of students (Cizek & Bunch, 2007).

Cizek and Bunch (2007, p. 6) believe standard-setting should not be left too late in the test administration cycle; it should start

> early enough to align with the identified purpose of the test; to align with the selected test item or task formats; when there is ample opportunity to identify relevant sources of evidence bearing on the categorical assignments; when that evidence can be systematically gathered and analyzed; and when the standards can meaningfully influence instruction, examinee preparation, and broad understanding of the criteria or levels of performance they represent. Indeed, the initial planning for any test that will be used to sort or classify individuals into performance categories should include a detailed discussion of the nature of those categories. A clear understanding of the performance categories will then influence all phases of test development, reporting, score interpretation, and validation efforts.

[84] Confusingly, In the USA the National Governors Association Center for Best Practices (NGA Center) and the Council for Chief State School Officers (CCSSO) have more recently used a very similar term 'standard<u>s</u>-setting' to refer to the process of writing the content standards in English Language Arts and Mathematics as part of the Common Core State Standards Initiative (NGA Centre & CCSSO, 2010).

Cut-scores define different categories, called performance levels, which can be labelled by *performance level labels* (PLLs) which name performance categories, for example, 'pass', 'fail', 'basic', 'proficient' and 'advanced', but do not inherently carry a description (Cizek & Bunch, 2007). *Performance level descriptions* (PLDs) provide a "fuller, more precise explanation of what the one-word PLLs attempt to convey" (Cizek & Bunch, 2007, p. 46), for example, the knowledge and skills which a candidate classified as advanced should exhibit. The PLDs "are supposed to anchor a spot on the continuum that is achieved by using the test to generate performance indicators … If the test is not tapping those skills and the PDLs are not constructed consistently with the test characteristics, you have a real problem" (R. Lissitz, personal communication, May 21, 2010).

Currently, the USA practice of standard-setting, especially concerning the validity of performance standards, is considered to be "perplexing" (McGinty, 2005, p. 270) to the psychometric measurement community because the "inner workings" (McGinty, 2005, p. 284) of judges and standard-setting cannot be discussed independently of a system of values (Messick, 1989b). Indeed, the "variety and demands of current contexts for standard[-]setting support our contention that the activity must be viewed as sophisticated and challenging" (Cizek & Bunch, 2007, p. 11). Consequently, there are a number of different models which are employed in standard-setting. These methods will not be reviewed in this thesis but the reader is directed to three excellent sources─Brennan (2006b), Downing and Haladyna (2006), and Cizek and Bunch (2007)─which describe and compare different standard-setting methodologies. It is important that whatever standard-setting method(s) be employed in decision-making, the results obtained be evaluated "using the same expectations and theoretical frameworks used to evaluate other measurement processes in education such as student measurement" (Nichols, Twing, Mueller & O'Malley, 2010, p. 14) because "cut[-]scores are constructed, not found" (Geisinger & McCormick, 2010, p. 38).

Science standards in the USA, both content standards and performance standards, were/are built by wide consensus (NRC, 2006) and standard-setting methodologies almost all involve some consensus of human judgments about student performances (Cizek & Bunch, 2007). In the USA definitions of content standards and performance standards remain distinct but interrelated with performance standards being derived from established content standards (Linn & Baker, 1995). This means that evidence of student performances, and of the assessment which resulted in those performances, have to be considered when making meaning of students marks.

The practice of standard-setting, including the determination of cut-scores, is different in the UK, in the Netherlands and in parts of Australia, and the practices for these countries are discussed below.

The ways in which standards are set in the USA, the UK, the Netherlands and parts of Australia all inform how standards are understood in this South African study (Section 3.5 below).

3.1.3 Standards in the United Kingdom, Australia [85] and the Netherlands

In the UK, the focus when determining educational standards is different to that described for the USA. Standard-setting, called awarding (Cresswell, 1998), is deliberately timed to occur once examinations have been written because it is related to only student performance in particular examinations rather than to the examination itself (Baird, 2000). The Assessment and Qualifications Alliance (AQA) (2007), the largest of the three English exam boards, stated that examiners try to set papers that are as uniform as possible in terms of difficulty between years; no explicit mention is made of content standards, or as to how the uniformity is maintained between years. This emphasis on examinees' performance rather than on the examination processes or systems has been highlighted by other researchers. For example, Cresswell (1998, 2000) believes that public examination procedures are designed for certification purposes rather than for systemic monitoring such as making comparisons of standards over time (Newton, 1997) or for determining a school's efficacy (Spencer, 2003). Similarly, Murphy (2003, p.189) stated that "[t]he prime function …[of examinations] is to award to individuals results which are as fair and reliable as possible year on year. It is not to monitor changes in standards over time". Greatorex (2003) summarized awarding as being concerned with *maintaining* a standard rather than *setting* a standard because the process was not concerned with *score interpretation* but with *score equating*.

A UK government body, the Qualifications and Curriculum Authority, has produced a Code of Practice (COP) which determines that two categories of materials must be used to set grade boundaries (cut-scores) in examinations: performance evidence and statistical evidence (Baird et al., 2000; AQA, 2005). The validity of such a standard-setting process depends less on the structure of examination question papers than on how students perform when writing the question paper. Researchers from The Associated Examining Board in England and the AQA (e.g., Newton, 1997; Cresswell, 1998; Baird, 2000; Baird et al. 2000) argued that educational standards are social constructs created by groups of judges (awarders) who are empowered by examining bodies and society to evaluate student performance in examinations. Similarly, Greatorex (2003, p. 6) of the University of Cambridge Local Examinations Syndicate reminded readers that "[w]hat is important is the reali[z]ation that standards [in the UK] are not objective measures, they are subjective judgements, or the consensus perception of senior examiners, or experts". None of these authors mentioned an explicit relationship between the question papers which resulted in particular student

[85] 'Australia' refers to specific Australian states, Queensland and New South Wales.

performances. Paterson (2003), in her explanation of standards, asserts that standards depend not only on what is assessed but also on the quality of the student responses, both of which involve human judgments of the people representing the examining body. It has been contended that if standards were considered as social constructs, they could not be predefined as they needed to be “dynamic and relative to specific moments in time” (Newton, 1997, p. 227) and can be culturally influenced (Cresswell, 2000). “Indeed, it [the social construct view] insists that standards do not exist at all, other than in the operational value judgments of those empowered to set them” (Baird et al., 2000, p. 227). Baird et al. (2000), in their study of the effect of coexisting definitions of educational standards on the fairness of examinations, argued against a single definition of examination standards for all examining bodies because the standards are subjective and might operate in different contexts, such as the different syllabi of the different examination bodies.

The various views of the concept of educational standards as a particular kind of social construct have been incorporated into assessments like the General Certificate of Secondary Education (GCSE) and A-level examinations in Britian (Cresswell, 1998; Paterson, 2003; Williamson, 2003). Examination question papers are set according to syllabus requirements for the different subjects. Once the examination scripts are marked, a group of experts—the awarders—in each subject considers the total marks awarded for the scripts in a particular examination. The awarders make their decisions about the grades[86] (similar to performance levels in the USA) that should be attached to particular marks based on their knowledge of previous students’ performance in similar examinations. The awarders also make use of archived scripts from the previous year’s examinations, examiners’ reports, and any other documentation and statistical information that might influence their judgment (Baird, 2000). In some instances, where examiners use “internally stored prototypical performances” to make judgments about standards, archival scripts are less important (Baird, 2000, p. 100). Awarding sets the specific grade boundaries (similar to cut-scores in USA)**,** including the passing grade for a subject (Cresswell, 1998). If analyses of the types of candidates sitting an examination indicated that the profile of typical candidates had changed, the statistically recommended grade boundaries could be changed to allow for any consequent effects associated with the change (Baird, 2000). In the UK context, ‘standard’ is used to define the performance required by examiners for specific grades in an examination, and to describe the general level of attainment of a cohort of students (Jones & Ratcliffe, 1996). Examination standards generated in the UK and the processes by which they are generated have been described as problematic, and examining boards could legitimately be called to defend their maintenance of such standards (Baird et al., 2000). Newton (2007, p. 7) described the approaches used in England for creating comparability by standard-setting as “quite fragile” because they were based on human

86 The number of examination grades used has been questioned (Cresswell, 1986).

judgment, there was little co-ordination between how different examining boards set standards, and, more importantly, because there was no consensus about how to achieve standards for comparability. Wiliam (1996a) recommended that the composition of the standard-setting panel be carefully considered and that the size of the panel be as large as can be properly managed, if year-on-year comparability within and between examinations were desired in the UK. Standard-setters making judgments about assessments should be full participants in the community of practice that they represent and trusted by the users of the assessment results, if the validity of the assessments is remain uncompromised (Wiliam, 1996b). As required by the COP, Newton (2007) recommended that approaches to awarding include both human judgmental methods, together with statistical methods.

French (1985) cautioned that the use of statistical analyses in understanding examinations should inform rather than mislead the examiners, and that statistical models be used not to rank candidates but to identify those candidates who should be drawn to the attention of the examiners, such as those who perform atypically. Once the examiners have awarded the grades, the statistical data could then be made available for *a posteriori* analyses (French, 1985). In the UK, statistical analyses are used to generate adjusted grade distributions (Jones, 1997), to compare the grading judgments of the awarders (Scharaschkin & Baird, 2000), the standards between years (Newton, 1997), the standards between different subjects (Goldstein & Cresswell, 1996) and the standards of different examining bodies that offer the same subject (Jones & Ratcliffe, 1996).

Across the world from the UK, two of the Australian states, namely, Queensland and New South Wales, have moved away from norm-referenced scaling of all subjects to standards-referenced reporting (Stanley & Tognolini, 2008). The use of standards-referenced reporting suggests emphasis on the importance of performance standards as criteria in criterion-referenced scaling. In a comparative study of top achievers in some subjects from four assessment bodies, two in the UK and two in Australia, which had standards-referenced reporting, Stanley and Tognolini (2008) suggested that the UK systems consistently had a higher percentage of top achievers than the Australian systems because of the manner in which the subject achievement level was used for university admissions. The Australian practice of statistically scaling subject performance to produce a university entrance ranking (Stanley & Tognolini, 2008) suggested that elements of norm-referencing were in operation. Tognolini and Stanley (2007, p 135) confirmed this view when they wrote "[to] arrive at realistic standards at some stage there must be some reference to normative expectations".

Tognolini and Stanley (2007, p. 129) further suggested that

> when constructing the standards for achievement with excellence, for example, the description could be written in such a way as to capture what it is that students in the top 10% of the subject would know and be able to do in relation to the particular achievement standard being considered.

This premise is based on an expectation that the top 10% of students all have a very similar performance, so the PLD would have to be relatively generic. "Semantic similiarity" in how performance standards are written, especially for the high achievers, makes it difficult to interpret the meanings of differences within subjects over years and between different subjects (Stanley & Tognolini, 2008, p.13). Given that PLDs need to be constructed consistently with the test characteristics (R. Lissitz, personal communication, 21 May 2010), one may question the value of the within year and between year comparisons between different examining bodies in the UK and Australia (Stanley & Tognolini, 2008) because in both countries comparisons were based on different examination question papers. Stanley and Tognolini's (2008) study made a tacit assumption─that the highest score in each of the examinations analysed would have the same meaning in terms of the content standards.

Elsewhere, in the Netherlands, the Dutch limited state control over curricular matters, and a flexible curriculum based on broad consensus in each subject was in place in schools and "the standards-setting process occurs through "strong democracy" and constant debate over what is worth knowing" (Louis & Versloot, 1996, p. 258). Prior to 1994, standard-setting at secondary school level was based mainly on adjusting the cut-score, between pass and fail, so that the same percentage of pupils passed the exams each year ─ which did not guarantee equivalence" (Alberts, 2001). Since 1994, in an attempt to address concerns about standards, the formal procedure of equating cut-scores,[87] using various different procedures has become part of examination processes (Alberts, 2001). Problems encountered during the equating of cut-scores in the Netherlands were, for example, the detection of discrepancies, between decisions made by different raters charged with deciding the cut-scores and changes in the performance of populations in different years, were such that "the distribution should not be taken (as it was) as a starting point for setting the norm" (Alberts, 2001, p. 365). Alberts (2001) concluded that the use of empirical evidence to inform decisions about cut-scores in the Netherlands needs to be ongoing as the education system evolves there. While maintaining standards in the Netherlands seemed to be focused particularly on performance standards (as in the UK) and norm-referencing, Alberts (2001) indicated that, because a rearrangement of the curriculum was giving greater emphasis to skills, there was the possibility of

[87] Called cut-off score by Alberts (2001).

some move towards the consideration of criterion referencing in future interpretations of examination scores.

The focus of standards in examinations in the UK-Australia-Netherlands group of examining authorities described here, is on the determination of grade boundaries or cut-scores, or what is known as standard-setting in the USA. While some implicit reference to content standards was made in the literature of these non-USA countries there was little explicit mention made of what performance standards might mean in terms of content standards. In the USA performance standards derive their meaning from the content standards.

Both the USA, and the UK-Australia-Netherlands groupings use specific, and somewhat different, views of standards to make inferences about a student's level of knowledge, skill or ability based on the students' test or examination performance. The degree to which these inferences are "correct" or defensible is called validity (Cizek & Bunch, 2007). In the USA, and its context of licensure and certification testing, validity depends on how the standards are set (American Educational Research Association [AERA], American Psychological Association [APA] and National Council on Measurement in Education [NCME], 1999), and on how the relationship between content standards and performance standards is articulated in practice (Cizek & Bunch, 2007). Validity of UK examinations would depend on the COP directive that two categories of materials, performance evidence and statistical evidence must be used to set grade boundaries (AQA, 2005; Baird et al., 2000). In the USA and in the UK-Australia-Netherlands standard-setting constructs new cut- scores each year, or for each test, in an attempt to ensure that performance level labels and their associated performance level descriptions have a similar or equivalent meaning between examinations and years. Therefore, studies of the examinations practice in the UK-Australia-Netherlands concern the understanding or meaning of performance standards, and therefore contribute towards our understanding of how performance standards are practiced in a South African context. This claim will especially apply to SC Biology answer scripts. However, studies from the UK-Australia-Netherlands offer no specific insight into how performance standards might be explicitly related to content standards, nor do they offer specific descriptions of performance standards as is available from studies of the USA.

3.1.4 Standards in South Africa

Goals of a scientifically literate citizenry, as well as good quality and equitable education, have been articulated for the people of South Africa for the 21st century (DoE, 2004e). However, there is no common understanding of what 'good quality' means, nor how or if good quality relates to standards or to the curriculum or to both, or even what standards are. Yet, articulated standards are "a basic prerequisite for the assessment of quality in education" Griffith (2008, p. 100). There was no published South African research literature which showed evidence of a specifically articulated common understanding of what standards mean in the local educational context. Some projects such as, for example, *Marking Matric* (Reddy, 2006a), brought together a number of scholarly works which all used the term 'standards' loosely and differently with respect to the SC examinations which take place at the end of high school.

End-of-secondary school examinations are one way in which countries operationalize what is important in their curriculum (Britton, 1996) and "[h]igh-stakes school examinations are one way that a society expresses the cognitive competencies that it values" (Genovese 2002, p. 101). Researchers such as Wiggins (1991a) and Steen (1999) argued that what we assess (i.e., examine) defines what we value, and consequently educational standards as practiced in examinations must be evidence of what we see as being important. If national secondary school-leaving examinations do reflect a country's educational standards (Eckstein & Noah, 1993), and tests (i.e., examinations) should be both standard-setting and standard-revealing (Wiggins, 1991b), then analysis of South African SC examinations should be able to inform our understanding of what educational standards are in South Africa at this level of schooling, (i.e., Grade 12). In South Africa, the SC examinations are school-leaving examinations and are high-stakes assessments because they determine whether students do or do not leave school with a SC qualification and whether students qualify to enter the job market and/or higher education. As high-stakes examinations, the SC examinations thus convey what the South African society values─its educational standards─when deciding if student is able to leave the secondary school system with or without this qualification.

South African researchers, for example, Zietsman and Gering (1986), Mitchell and Fridjhon (1987), Mitchell, Fridjhon and Haupt (1997), Griesel (2001), Yeld (2001), Fraser and Killen (2003), and Naidoo (2006), have all questioned the predictive value, and therefore the standards, of SC examinations for both university admission and success, and by implication have therefore questioned the validity of these examinations. The concern about questionable value of secondary school exit qualifications for entry to higher education is not unique to South Africa (Bakker & Wolf, 2001; Harman, 1994; wa Kivilu, 2006). In South Africa, the lack of performance of students at higher education levels could be the result of either a lack of "synchronization" between what secondary school offers and what higher education requires, or the possibility that "standards" in secondary schooling are falling (wa Kivilu, 2006, p. 36).

The nature and extent of any evidence provided by the South African SC examinations, and student performance therein, of standards (or what we valued) is unstudied and elusive. In South Africa, it was not clear what “counts as worthy work to be mastered” (Wiggins, 1997, p. 56) (i.e., content standards), or what different levels of mastery mean (i.e., performance standards). Mastery, *sensu lato*, is that of Bloom, Hastings and Madaus (1971, p. 56) “[mastery] must be … a public recognition by school or society” and in this thesis is used synonymously with competency. The public recognition must be in the form of appropriate certification by the teacher or the school”.[88] Hiebert (1999) considered education standards to be statements about purposes, priorities and goals. In one of the subjects offered at SC level, Biology, there is little or no documentation available, other than syllabi and guideline documents from which expectations of pupils at the SC level could be inferred. While the objectives and goals are embedded in the CBS (see Chapter 4) they are not explicated in the same way as in USA science standards documents. Therefore, in South Africa the educational expectations or standards were not explicitly articulated nor was the process by which standards were understood from examinations explicitly conveyed to society, in contrast to, for example, the USA and the UK. In South Africa, the cut-scores separating different levels of mastery remained fixed between years and were the same for all subjects (JMB, 1989), that is, they were not set independently for each examination, based on student performance therein as they are in standard-setting processes of the USA and the UK-Australia-Netherlands. The equivalence of marks awarded between different subjects and Grades in South African SC examinations has been challenged (Oberholzer,1995). Thus the challenge for this thesis was to find a way of explicating the implicit standards of the SC Biology examinations for the examination question papers and from candidates’ answer scripts.

Provided standards are made explicit in an education system, and that they are sound,[89] they have the potential to improve our understanding of the relationships between student performance and the other components of the curriculum (Porter & Smithson, 2001) (see Section 3.1.2), “predicated on the assumption that what is taught is a strong predictor of what students will learn” (Porter, Polikoff & Smithson, 2009, p.238). It is for this reason that what it means to have a standards-based curriculum is explored further below.

3.1.5 What it means to have a standards-based curriculum

A review of the literature about education systems across world showed that the term *curriculum* can have different meanings and is used in very different ways, both within and between countries,

88 Bloom (1968) in discussing *learning for mastery* proposed “mastery” to mean absolute or 100% learning of all the objectives of instruction which is slightly different to how the term is used in this thesis, which recognizes levels of mastery.

89 The NRC (2012) considered the Framework for K-12 Science Education in the USA to be the critical first stage in developing the revised Science Standards because it was rooted in current research on science and science learning, and it identified the science that all students need to know.

to describe different components of a particular country's education system. For example, in some countries curriculum means simply the topics to be studied or the guidelines for courses of study (Schmidt et al., 2005; Valverde, 2000, 2005). Section 3.1.2 described current practices in the USA where standards are considered to drive the curriculum. This section will trace the evolution of a particular view of curriculum that embraces much of what happens in a school or education system. In doing so, it illustrates the framework that standards can bring to understanding the relationships between components of a curriculum and student performance.

Preceding the advent of a standards-based curriculum in the USA, Travers, Crosswhite, Dossey, Swafford, McKnight and Cooney (1985) described the curriculum as tripartite in nature, for their analyses of data concerning student performance in the Second International Mathematics Study (SIMS). The three components of curriculum recognized by Travers et al. (1985) were the intended curriculum (i.e., the intent of education system), the implemented curriculum (i.e., the instructional practice), and the attained or learned curriculum (i.e., what students have learned). These authors conceptualized a linear relationship between their intended curriculum, implemented curriculum and attained curriculum (Figure 3.3). When Travers et al. (1985, p. 3) described the intended curriculum as being "reflected in curriculum guides, course outlines, syllabi and textbooks adopted", they were describing the curriculum component that would become the 'standards' component within the standards-based curriculum of the future. Compartmentalizing curriculum as in Travers et al. (1985) continues to be a strategy to describe curriculum in contemporary USA literature, but the names of the curricula components vary across contexts. For example, in science education the term intended curriculum refers to official documents like the published science standards, while the implemented curriculum is also known as OTL (McDonnell, 1995; Porter, 1995) or the enacted curriculum (classroom instruction) (Porter & Smithson, 2001), and the attained curriculum is known as the learned curriculum (student outcomes) (Porter & Smithson, 2001). Porter and Smithson (2001), added a fourth component to those previously recognized by Travers et al. (1985), called the assessed curriculum, to distinguish assessments from the rest of the curriculum. Porter and Smithson (2001, p. 2) suggested that although the learned curriculum might be considered an aspect of the enacted curriculum, it was wiser to keep them separate because "these finer distinctions serve an important analytic role in tracing the chain of causality from education legislation [and input variables] to student outcomes [output]".

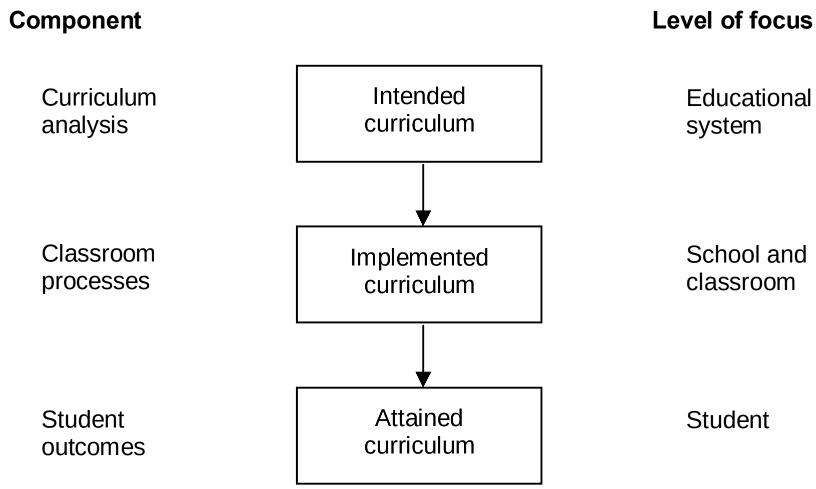


Figure 3.3 The SIMS conceptual design of curriculum (Travers et al., 1985).

Porter and Smithson (2001) also suggested other relationships, for example, between the intended curriculum and the attained curriculum and their additional component, the assessed curriculum. This suggestion lead to a more complex structure of curriculum (Figure 3.4) than that described by Travers et al. (1985). The triangular and reciprocal relationships between curriculum (intended curriculum),[90] instruction (enacted curriculum) and assessment (assessed curriculum) illustrated in the Porter and Smithson (2001) model were first described by Tyler (1949). Tyler's view of curriculum was criticized because of claims that its simplicity obscured the complicated issues and problems behind both curriculum design and curriculum development (Kleiband, 1970) and it lost favor. Later, a colleague of Tyler, Benjamin Bloom reiterated the importance of the inter-connectedness of these components in educational programs (systems) and also recognized a component which described "what students actually learn" (Bloom, 1975, p. 11, emphasis in original) (learned curriculum).

In positioning of the learned curriculum (i.e., the attained curriculum of Travers et al. [1985] or student achievement) at the centre of their model of curriculum, Porter and Smithson (2001) signaled the importance of recognizing that the learned curriculum is a combined expression of the interplay between the intended curriculum, the enacted curriculum and the assessed curriculum. This view is consistent with how Finn (1990) conceptualized a curriculum. Finn (1990) argued that education should be defined in "terms of learning" (p. 589) or "in terms of ends rather than means" (p. 581) and hence a curriculum should be considered to be output-based, where outputs were defined in terms of specific student learning. Student learning is an undisputed goal education. Other models of curricula, for example, Schmidt, McKnight, Valverde, Houang and Wiley (1997)

90 Bracketed terms in this paragraph were not used by Tyler (1949) or by Bloom (1975) but are given to indicate conceptually comparable terms used by Porter and Smithson (2001).

have linked student performance (outcomes or achievement) more closely to the assessed curriculum than to the intended curriculum.

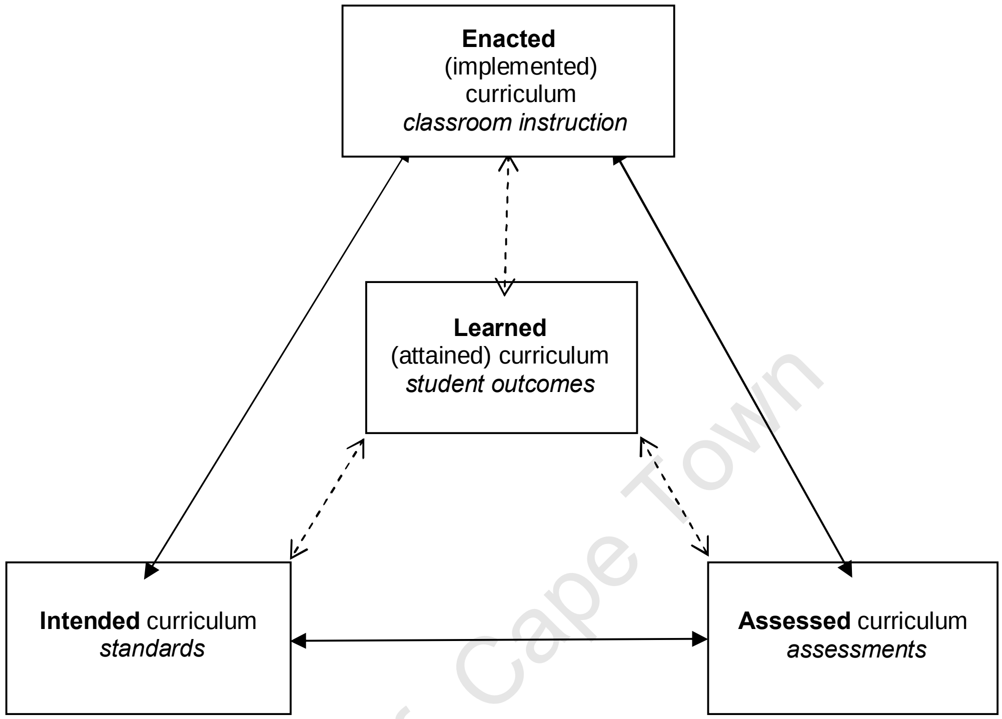


Figure 3.4 The relationship between curricula components showing alignment analyses (─) and proposed alignment analyses (---) conducted in the USA (modified from Porter & Smithson [2001])

"Distinguishing the four components of the curriculum delivery system allows for examination and comparison of the curriculum at different points [components] of the system" (Porter & Smithson, 2001). Although Porter and Smithson described their curriculum model as a standards-based system, aspects of the model and its precursor have been successfully used for studies of the curricula of countries without obviously articulated standards as in parts of the USA, where the intended curriculum was taken from other documents such as syllabuses and guidelines. For example, the SIMS (Travers et al., 1985) and TIMSS (Schmidt et al., 2005; Valverde, 2000) projects examined relationships between the intended curriculum, the enacted curriculum and the attained curriculum in an attempt to understand how variation in the intended curriculum and the enacted curriculum of different countries related to variation in the attained curriculum as measured by student performance.[91] Porter and Smithson (2001) termed such studies 'curriculum indicators' because they were able to 'tell' investigators something about the curriculum. Curriculum

91 Because a nation's curriculum is an enactment, through policy and practice, of its intended educational aims [standards] and "its vision of itself and its place in the community of nations" (Schmidt &

indicators are useful in that they are able to describe the educational opportunities in schools, monitor reform and explain student achievement, or its absence (Porter, 1991; Porter & Smithson, 2001) .

Developing the instruments and methods to generate curriculum indicators is considered to be complex (Porter & Smithson, 2001) and dependent largely on the potential use of the curriculum indicators (Porter, 1991). Curriculum indicators are less complex in a standards-based curriculum, and particularly in the Porter and Smithson model of curriculum (Figure 3.4), because the standards are common to all curriculum components. One curriculum indicator described by Porter and Smithson (2001) was *alignment* which assesses the 'match' between two selected components of the system based on comparable descriptions of each component. The concept alignment is discussed separately in Section 3.3 below. For the purposes of the discussion which follows in this section it is sufficient to understand alignment to be a measure of the match between any two curricula components. In an ideal standards-based curriculum, the standards provide the framework or "common language" against which the comparisons are made, provided that the education system has the standards properly articulated in the first place. In their work, Porter and Smithson (2001) described a number of successful alignment studies which have been carried out in the USA as well as potential alignment analyses which may be carried out in pursuit of "a common and systematic language for describing key elements of the curriculum" (p. 21).

The Porter and Smithson (2001) model of curriculum is in contrast to other, more complicated, input-output process driven models (e.g., Poliah, 2009; Porter, 1991; Schmidt et al., 1997; Taylor, Muller and Vinjevold, 2003) which link student achievement only indirectly to the input of the systems through a 'black box' of processes which have to be navigated when trying to understand student performance. Consequently, the Porter and Smithson model of curriculum can be used as a framework to understand the functioning of education systems at various levels, such as, that of the school, district, province or nation, depending on what question is being asked about a particular educational delivery system (Porter & Smithson, 2001).

In an ideal education system if teachers teach what the standards say they should teach, and assessments are valid because they assess the same standards, students would have been taught what is in their assessments and therefore students are prepared for their assessment. Instructional approaches would be aligned with *both* the content and the performance standards (NRC, 2006). Teachers believe that published standards are sufficient to have resulted in improved learning (DeRoche, 2005) and increased student achievement and learning (Swanson, 2006). Other studies

McKnight, 1995, p. 346) inter-curricular comparisons, like TIMSS, within and between countries, offers insights that intra-curricular studies within a country might not reveal.

have shown that standards-based education systems whose components are all aligned as described above are rare (Bybee, 2009). For example, problems arise when what is required of students (standards) may be set too low, and teachers teach at this level (Janofsky, 2005; Thomas B. Fordham Foundation, 2005). Despite this, Feuer (2010) and Huff and Plake (2010, p.141) cautioned against letting "our work towards the *optimal* get in the way of *reasonably good*" when trying to design curriculum components such as assessment systems. Regular critical evaluations, particularly with respect to both the benefits and downside risks associated with how student achievement is interpreted and used might bring us "reasonably close" to a set of "desirable conditions" in an assessment system (Feuer, 2010) and therefore our understanding of the dynamics of an education system.

Alignment between standards and assessments has been used as an accountability criterion (Bricker, 2002; La Marca, 2001). Theoretically, standards in a standards-based education system should bring equity to school systems by providing an equal access to all students but practically this equity does not guarantee equality of outcomes because of conflicting interests (i.e., misalignment) between components of education systems (Darling-Hammond, 2007; Robinson, 2011). Too often stakeholders in an education system, such as policy makers, intended curriculum specialists and teachers, are distrustful of one another and therefore function relatively independently of one another (Popham, 2004)─a condition not conductive to constructing an aligned standards-based system as described above. Potterton (2006) argued that standards can do little to change teaching and learning in contexts of poverty such as in parts of South Africa.

Teaching-to-the-test could occur, for example, if the same or very similar tests are used repeatedly within the system, or when guidelines other than the official policy documents which define the standards to be assessed are circulating within the system. Examples of such documents are exemplars used for the professional development of teachers or for the coaching of students.[92] Alignment between the standards and the teaching would indicate that teachers are only teaching a subset of the intended curriculum, namely, the subset that is tested. Student learning might appear optimal because the students do well. In this scenario, even though alignment analyses between the standards and an assessment may appear to be acceptable (bearing in mind that a test can only ever be a subset of the domain covered by standards), a thorough validation process should detect that the validity of the examination is compromised because of, for example, the assessments being highly predictable. Although alignment is a required condition of validity evidence in a standards-based systems such as in the USA (NRC, 2006), it does not necessarily guarantee validity.

[92] The practice of using exemplars in this way has become practice in South Africa at SC level in recent years.

Teachers may teach-to-the-test because the standards are vague, not well written or incomplete. In such a scenario, the tests become the *de facto* standards which is not what was originally intended by the system. Checking alignment between the standards and the teaching should indicate that teachers are only teaching a subset of the intended curriculum, that is, the subset that is tested. If references between instruction and assessment are not made back to the intended standards, and to the strength of the alignment, the misalignment in such a curriculum might be missed.

Two examples of teachers teaching-to-the test rather than to the standards comes from the USA. Science teaching was found to be directly influenced by state tests and assessment anchors, which varied significantly over time, rather than conform to the state standards, which were more stable over time and which may or may not have been influenced by national standards (Schunn, 2009). Therefore, varying tests, even if they were all valid, must have sent different messages to teachers at different time times as to what was important to teach whereas a focus on the more stable state standards would have sent a consistent message (Schunn, 2009). In another study, Shepard (2010) reported that because most formative and interim assessments in the USA were structured to prepare students for high-stakes summative assessments (many of which did not often reflect the standards, and were therefore not valid) the assessments lost their capacity to be used diagnostically to improve teaching and learning.

Therefore, standards are a reference point against which both qualitative and quantitative evaluations of the various components of a curriculum are to be made (NRC, 2006). However, to be able to function as a reference point for comparisons, standards need to be articulated in specific ways. The TIMSS analyses found that the countries with the highest level of mean student achievement on the mathematics and science tests generally had intended curricula that were like standards, that is, prescriptive (Schmidt et al., 2005). However, it is not enough simply to have content standards − the content standards needed to be of a good quality and coherent. Standards should

> specify topics, including the depth at which the topic is to be studied as well as the sequencing of the topics, both within each grade and across the grades, in a way that is consistent with the structure of the underlying discipline. (Schmidt et al., 2005, p. 554)

Coherence of content (standards) is critical for teaching and learning for understanding, as argued Schmidt et al. (2005, p. 554):

> Understanding implies, at least at some level, that the structure of the discipline has become visible to the learner so that she or he can move beyond its particulars. We suggest that one way to facilitate such learning is by making the inherent logical structure of the discipline more visible to both teachers and students [through standards].

Moss (2004) questioned the widely accepted premise that coherence within school systems is necessarily a desirable objective. Using sociocultural theory, Moss argued that even when standards, curriculum, instructional practices, and assessments together represented sound pedagogy, important learning outcomes might still be compromised. If an education system expects the same set of learning outcomes of all teachers and all students, diversity in learning experiences will be narrowed. Moss (2004, p. 233) proposed a model of "coherence-through-negotiation-of-meaning" which recognizes diversity as a resource for learning and for social change and which allows the flexibility for adjustments to be made to common standards based on unique contextual circumstances. However, Rothman (2009, p. 5) cautioned that standards which arose "by consensus, rather than vision" have often resulted in a tendency to overload the content standards rather than focusing on the precise content and skills students are required to learn.

"A policy is prescriptive when it is explicit in describing what content decisions are required" (Porter, Floden & Fuhrman 1998, p. 130). Therefore, in addition to being coherent, standards should be explicated in unambiguous language in order to influence other components of the curriculum. Coherent content standards written in unambiguous language should lean towards strong policies which are the directives by which the components of curriculum operate. Hill (2001, p. 289) gave the example, "[w]here state standards used words like "construct" and "concept" to imply certain mathematics teaching methods, teachers reading those documents imputed more local, and sometimes conventional, definitions to these words. As a result, state standards lost their force". Here, the lack of commonly understood language was a barrier to implementation of the standards.

Porter et al. (1998) gave another policy characteristic for standards: consistency. When policies are consistent, they send the same message to all components of the education system. Hill (2001) suggested that in addition to simply directing an education system or curriculum as to what to teach, to learn and to assess, the policy (standards) should have a more educative function by explaining the decisions of policymakers (i.e., writers of the standards), decisions in terms of reasons why they included the standards that they did.

In the absence of well explicated standards, textbooks (Schmidt et al., 2005) and assessments become *de facto* standards. In effect, this inference means that even when countries, like South Africa, do not necessarily think of their syllabuses and guideline documents as standards because the information they contain is not conveyed with that claim, these documents can yield some information about curricula in a similar way that standards in a standards-based-curriculum do. The difference is that in a standards-based-curriculum, (content) standards 'drive' all of the other curricula components so that textbooks, syllabus and guideline documents do not necessarily guide the curriculum in the same way as without the standards.

However standards are defined by and education system they give a reference point or common language by which society can try to make sense of how an education system is functioning. Standards can fill this role because provided they are explicit or can be made explicit, they make it consistently clear to each part of the system what is required of it.[93] Standards therefore potentially bring equity and inclusivity to education systems (McClure, 2005) as required by law in many countries, including South Africa. Despite the theoretical promises of the advantages of standards for a curriculum, implementation in the USA has been far from perfect with respect to equity. For example, Darling-Hammond (2003, p.1) stated that in the USA, while standards-based reform had been partly successful in improving the quality of education in some states, if standards and assessment were "to support improved education rather than greater inequality" and to improve student learning, then the quality and alignment of standards, curriculum guidance and assessments would need to be addressed. Darling-Hammond explained that in the USA, despite having standards, the gap was becoming greater between teaching and learning in some groups of children because of a "collision of new standards with old inequalities" (Darling-Hammond, 2007, p. 318). Darling-Hammond proposed reforms which equalized opportunities to learn for all students as a necessary condition for standards to succeed in curriculum reform (Darling-Hammond, 2007). It has been argued that while standards might drive testing as a proxy for equity, they do not change the underlying education system. Instead equity could be more ensured by teachers in the classroom than by standards (Richardson 2009/2010). Darling-Hammond, Richardson and the NRC (2006) highlighted the need to remember the importance of the education system as a whole, and the role of all the stakeholders in practicing the standards, when examining the realization of the intended curriculum or standards.

A recent Brown Center Report on American Education questioned the value of a standard-based curriculum to improve student learning: "Standards in education are best understood as aspirational ... they represent good intentions not often realized" (Loveless, 2012, p. 13). Reasons given for why in the USA standards had little effect on changing schools and what students learned are the enormous variation in OTL within schools, within states and between states because the "proper role of federal government, states, local districts and schools in deciding key educational questions especially in deciding what should be taught, remains a longstanding point of dispute (Loveless, 2012, p. 13-14). Despite Loveless' misgivings about standards improving learning in an education

[93] Bernstein (1977) identified three message-systems in education systems – curriculum, pedagogy and evaluation. Ideally, in a standards-based curriculum as described here, there is one consistent 'message', the standards, which operate in the school system.

systems, standards have other educational functions. For example, provided standards are explicit, they provide a consistent language or framework to trace “chain of causality from education legislation to student outcomes” (Porter & Smithson, 2001, p. 2), and help to understand education systems. Provided that the standards represent the concept domain of a discipline (Yager & McCormack, 1989), they can promote conceptually sound models of teaching and assessment (Enger & Yager, 2001).

Others, such as, for example, Huff and Plake (2010), described a standards-based education system as a “teaching and learning ecosystem” (p. 140) with a “diversity of actors … who continually influence this system, the interdependency of the various elements, and the iterative, dynamic nature in which these various components change and evolve” (p.141).[94] This view of Huff and Plake is consistent with the view of standards constructed for this thesis (Section 3.1.1) ─the actors are the “stakeholders” and “recognized authority” and the notion of an evolving education system is captured by the terms “reviewed regularly and revised if necessary”. Huff and Plake (2010, p. 142) reiterated the importance of “a robust articulation of the domain of student learning”, that is, the content and performance standards are at the “core of a healthy teaching and learning system”─which is consistent with the model of Porter and Smithson (2001).

In an explicitly standards-based curriculum, the intended curriculum (the standards) is supposed to guide both what is taught (the enacted curriculum) and what is tested (the assessed curriculum). Validation is about making meaning of test scores or student performance (Cizek et al., 2008; Cizek et al., 2010; Killen, 2003; NRC, 2006). In a standards-based system, the validation of assessments would involve determining whether the assessments adequately represent the standards. Therefore, in a standards-based-curriculum, the standards give meaning to scores and provide a benchmark for comparisons of test scores over time. The validity evidence generated in the validation of assessments should thus mirror (or not mirror) the standards.

In South Africa, SC Biology was not part of a standards-based curriculum which meant that the standards of the SC examinations were implicit rather than explicit. Therefore, if comparisons between the standards of the SC Biology examinations are to be made in this thesis (Chapter 1), content standards and performance standards had to be made explicit from those documents in which they were implicit, that is, the CBS and versions of the CBS which guided the SC Biology

[94] Huff and Plake (2010) were particularly concerned that understanding student performances in assessments was being hampered by the limitations of methodologies by which performance standards were set. Despite using slightly different terminology for the four components recognized in Figure 3.2, the components described by Huff and Plake (2010) are conceptually similar. For example, these authors refer to the learned curriculum in terms of learning progressions described by performance level descriptions which describe the performance of students in assessments.

examinations, SC Biology question papers and student answer scripts. If validity evidence is a mirror for explicit standards in a standards-based curriculum, then presumed validity evidence should potentially be a proxy for implicit standards. To test the potential of validity evidence as a proxy to explicate the standards implicit in the SC Biology examinations, it was necessary to explore the role of validity in assessment.

3.2 Validity and assessment

Validity and reliability are important for all assessments (Brennan, 2006a; Downing & Haladyna, 2006), including those that take place in the classroom during the year (formative), the more formal assessments like tests and examinations (summative) (Popham, 2009; Shepard, 2003), and those that are conducted throughout the year and mirror summative assessments (interim) (Popham, 2009). The purpose or use of an assessment influences the way it is designed and how its validity and reliability are evaluated (Shepard, 2003). The assessment literature indicates confusion about how validity and reliability are both viewed and used, and about their relationship to one another (Lissitz, 2009). In particular, while the measurement literature tends to separate reliability and validity, the terms are sometimes taken as being equivalent, or reliability is taken as evidence of validity (Lissitz, 2009). Given that "[t]he concern of reliability is to quantify the precision of test scores and other measurements" (Haertel, 2006, p. 65), reliability of assessments is subsumed into the discussions of validity which follows. Specific issues related to reliability, particularly, the issue of reliability of the SC Biology examinations scores is addressed in Chapter 5. The issue of reliability of the classifications of the cognitive demand of the SC Biology examination questions is discussed in Chapter 4. Indeed, validity can also be a confused concept (Gorin, 2007; Lissitz, 2009; Sireci, 2007), because of a mismatch between the theoretical aspects and the practical applications of validity (Lissitz, 2009).

What follows below is a summary and discussion of some of the validity literature which the author has deemed necessary to explain how she came to use modern conceptualizations of validity to explicate standards of SC Biology examinations in the conceptual framework developed in this chapter (Section 3.5.1). The USA is the only country known to the author that has associations of psychologists and educationalists which jointly provide guidance about the validation of assessments through regularly updated versions of *Standards for educational and psychological tests/testing* (APA, 1954; APA, AERA, & NCME, 1974; AERA, APA, & NCME, 1985, 1999). Consequently, most of the literature on the validation of assessments in this thesis comes from the USA.

3.2.1 The evolution of the concept validity as applied to assessment

Validity assertion as a part of test practice has been documented for a long time. In his brief history of validity theory, Sireci (2009) described some of the first forms of validity theory and test validation that were recorded in conjunction with the advent of tests in the late 1890s. In 1896, Pearson developed a correlation coefficient, "an exciting new statistical index" (Sireci, 2009, p. 21) which allowed validity of a test to be defined by a correlation between a test score and some criterion considered to be important to the meaning of the score. This approach to validity has been termed the criterion-related validity model (Haladyna, 2006; Sireci, 2009). Although this approach to validity is still used in admissions testing, employment testing and credentialing testing (Haladyna, 2006), it has been criticized because the criterion chosen are "not adequate" (Kane, 2006a, p. 18), or are "imperfect at best" (Sireci, 2009, p. 23) or were not able to be validated (Kane, 2006b). These limitations could be addressed if testing programs, that make a prediction or a correlation to a criterion, document the consistency of the decisions by which the prediction or correlation was made (Haladyna, 2006).

In the 1930s, another statistical method, factor analysis, as developed by Spearman in 1904, emerged as a mechanism of representing latent psychological 'traits' or 'constructs' measured by tests (Sireci, 2009). Around about the same time, a theoretical definition for validity as "*the degree to which "a test measures what it is supposed to measure"* (Garrett, 1937, p. 324)" was proposed (Sireci, 2009, p. 22). Dissatisfaction with this second theoretical definition of validity together with assertions that tests could be validated by using only statistical methods, appeared in the 1940s (Sireci, 2009). Psychometricians began "to consider more carefully the attributes intended to be measured by a test as well as those reflected in criterion performance" (Sireci, 2009, p. 23) and to find ways to operationally define the attributes that had to be measured. What emerged at this period has become known as the content validity model because it involved analyses of test content together with the criteria with which the tests were meant to be correlated (Sireci, 2009). "Content validity is established by showing that the test items are a sample of a universe in which the investigator is interested" (Cronbach and Meehl, 1955, p. 282). Focusing on content validity in this way was considered to be limiting as it "tends to be subjective and to have a confirmatory bias" (Kane, 2006a, p. 19) and because it did not involve the consideration of student performances on the test (Messick, 1989a). Content-related evidence can be used to justify the criterion measures in the criterion related model (Kane, 2006b).

Psychometricians had still not reached consensus about validity theory and test validation (Sireci, 1998, 2009) so the *Technical recommendations for psychological tests and diagnostic techniques*

(APA, 1954) was written to guide test developers and test users. In this article, four types or attributes of validity were identified: concurrent validity, predictive validity, content validity and construct validity. "The first two of these [concurrent and predictive validity] may be considered together as *criterion-oriented* validation procedures" (Cronbach & Meehl, 1955, p. 281, emphasis in original). *"Construct validation* is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not "operationally defined." The problem faced by the investigator is, "What constructs account for variance in test performance?" (Cronbach & Meehl, 1955, p. 281, 282, emphasis in original).

Content validity was considered more important for educational testing (Sireci, 2009) and Cronbach and Meehl (1955) proposed that construct validity be used as an alternative to criterion and content validity especially for psychological tests. Loevinger (1957, p. 636) supported this view "since predictive, concurrent and content validities are all essentially *ad hoc*, construct validity is the whole of validity for a scientific point of view". Much later in the evolution of validity theory, Cronbach , Meehl and Loevinger's view of validity was embraced by the unitary conceptualization of validity where all validity was considered to be construct validity (Messick, 1989a)─a conceptualization of validity that remains popular today (Sireci, 2009) although theoretically it is not considered to provide sufficient guidance for applied work (Lissitz, 2009). The notion of an holistic view of validity is supported in the *Standards for educational and psychological tests* (APA, AERA, & NCME, 1974) and again in the *Standards for educational and psychological testing* (AERA, APA, & NCME, 1985).

Kane (2006a) summarized three principles which emerged out of the construct-validation model (Cronbach & Meehl, 1955; Loevinger, 1957; Messick, 1989a) which were still operating in the 1980s. These principles were that the proposed interpretation to be made from test scores be declared before its validity is evaluated; that multiple sources of evidence rather than a single source of evidence be used; and that both the proposed interpretation and competing interpretations be considered. The meanings of construct validity have continued to evolve with time and continue to be used in different ways (Zumbo, 2009) and in confusing ways (Sireci, 2007). Some authors, for example, Borsboom, Cramer, Kievit, Scholten and Franić (2009), recommended that construct validity be abandoned because of the disjunction between how validity theoreticians present validity and how practitioners practice validity.

Chronologically the development of the argument-based approach to validation followed the construct validity model. Cronbach (1988) proposed that the logic of a validity argument be used to evaluate the inferences and decisions being made from test scores using validity evidence. The term "validation" and to a lesser extent the term "validity" are used in two distinct but closely related ways in discussions of measurement. For example, "validation" involves the development of evidence to support the proposed interpretations and uses; and "to validate an interpretation or use" is to show that the interpretation is justified (Kane, 2006a, p. 18). Another example is the use of "validate" to refer to the extent to which the proposed interpretations of test data are plausible and appropriate (Kane, 2006a). The argument-based approach to validity might be described as "a compromise between sophisticated validity theory and the reality that at some point, we must make a judgment about the defensibility and suitability of a test for a particular purpose" (Sireci, 2009, p. 29). Validity argument is as strong as its weakest link (Cohen, Kane& Crooks, 1999).

Messick (1989a, p. 13) defined validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales [which] support the adequacy and appropriateness of inferences and actions, base [themselves] on test scores or other modes of assessment". Kane (1992, p. 527) combined the 'evaluation' and 'validity argument' components of Cronbach's vision and Messick's definition and termed validity <u>an evaluative (validity) argument</u> where

> [v]alidity is associated with the interpretation assigned to test scores rather than with the test scores or the test. The interpretation involves an argument leading from the scores to score-based statements or decisions, and the validity of the interpretation depends on the plausibility of this interpretive argument.

Kane (1992, p. 534) described the interpretive evaluative argument as an "approach to validity rather than a type of validity" which avoided the compartmentalized view of validity as the categories concurrent validity, predictive validity, content validity or construct validity. Kane's view of validity as an evaluative argument focused empirical investigations on the most vital issues, was supported by others (e.g., Shepard, 2003) because it directed research effort into identifying the kinds of validity evidence required (Cronbach, 1988) and attempted to bridge the gap between theoretical validity and practical validity studies. Kane (2001, 2002) expanded his earlier focus by emphasizing the role of validity evidence in the refinement of the interpretive argument until the validity becomes defensible or until it is rejected. In the argument-based approach to validation, validity evidence is presented in a robust and substantiated way (Kane, 2001).

In more recent work on the argument-based approach to validity, Kane (2006a, p. 22) identified two separate parts to validation, "the specification (the interpretive argument) and evaluation (the validity argument) of the proposed interpretations and use of the scores". The interpretive argument explicates the "inferences and assumptions inherent in the proposed interpretations and uses of these results" and the validity argument "requires the integration of different kinds of evidence from different sources" to support or challenge the proposed interpretations (Kane, 2006a, p. 23), and to refine the interpretive argument if and/or where necessary. The interpretive argument also helps to "protect against inappropriate interpretations and uses [of test scores] by making gaps in the evidence harder to ignore" (Moss et al., 2006, p. 117). The argument-based approach to validity therefore helps to identify the appropriate kinds of validity evidence required to address specific inferences and assumptions made about test scores (Kane, 2006a). In high stakes testing, the proposed interpretations and the methods by which tests scores are used to make decisions might need to be negotiated by stakeholders or users of the test (Ryan, 2002). The choice of validation argument would depends on both the context and the 'metaphor' (an understanding the psychometric properties) that drives the test (Mislevy, 2009).

The current version of the *Standards* (AERA et al., 1999, p.17) promotes the argument-based approach to validation as "[a] sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of test scores for specific uses".

In two articles on validity, Kane (2006a, 2006b) listed the criteria for evaluating interpretive arguments in slightly different ways. These four criteria which are detailed below because they are particularly useful as a framework for interpreting test scores.

1. Clarity of the argument − the interpretive argument should be clearly stated as a framework for the validation (Kane 2006a, p. 29). The inferences and assumptions used to move from test scores to the proposed conclusions and decisions should be explicitly stated (Kane, 2006b, p. 139).
2. Coherence of the argument − the interpretive argument should be coherent in the sense that the conclusions should follow from the test scores and the supporting assumptions. Interpretive arguments are informal or presumptive arguments which are not expected to provide a logical or mathematical proof of any conclusions, but rather to establish a strong presumption in favor of the proposed interpretations and uses of test scores (Kane, 2006b, p. 139). The argument is expected to be complete in the sense that no essential inferences or assumptions are omitted (Kane, 2006a, p. 29).
3. Plausibility of inferences and assumptions − the assumptions included in the interpretive argument should be plausible. While some assumptions may be established well enough to be taken for granted (Kane, 2006b, p. 139), other assumptions can be supported by various

kinds of evidence, for example, careful documentation and analysis of procedures and empirical evidence (Kane, 2006a, p. 29).

4. Evaluating other interpretations − any plausible alternative interpretations of test scores, including construct-irrelevant variance,[95] should be identified and eliminated if possible (Kane 2006b, p. 139). The plausibility of an assumption is judged in terms of all the evidence for and against it (Kane, 2006a, p. 29). One of the most effective ways to challenge an interpretive argument is to propose an alternative argument that is deemed more plausible (Kane, 2006b. p. 139)

Both Brennan (2001) and Haertel (1999) cautioned against the danger of collecting pieces of evidence using a checklist approach, as in a framework, while ignoring how the pieces fit together. Kane's evaluative validity argument view of validity described above is consistent with the view of Messick (1989a) because it focuses attention explicitly on the chain of inferences which link the specific performances on a particular test to the "target domain" about which we want to make decisions and draw conclusions (Moss et al., 2006, p. 117). The evaluative validity argument requires that the validity evidence fit together in some logical way. How the target domain is defined then becomes important. It has been suggested that defining the target domain of particular validation processes might be achieved by asking questions related to the validation process. For example, Brennan (2001, p. 12) warned against assuming that an "unordered accumulation of vaguely focused studies constitutes validation" or the other extreme which considers "a single unifying, integrated framework to encompass all of validation". Brennan (2001) proposed that researchers use what he called "Socratic validation". The Socratic validation approach that Brennan described requires that clear, pertinent questions be asked, and that these questions be answered by interrogating the appropriate evidence. Similarly, Moss et al. (2006) called for a more flexible approach to validity that starts with relevant questions being asked of the validation process. Haertel (1999) stressed how important it was that the correct questions be asked in validation studies.

Brennan (2001) gave as examples some important questions that could be asked, especially those concerning content validity and the statistical specifications of a test, as without these elements test scores cannot be meaningfully interpreted. "In Socratic validation, or any approach to validation, for that matter, we must be prepared to acknowledge that some of our questions will not be answered and/or some of our answers will be less than flattering" (Brennan, 2001, p.13). Using a

[95] In high-stakes examinations, scoring errors included in the final test scores reduce validity evidence and the credibility of the examination because such errors introduce construct-irrelevant variance to the scores (Haladyna & Downing, 2004).

Socratic validation approach to validity would therefore also minimizes the possibility of the confirmationist bias described by Cronbach (1988), which might lead us to ignore "actively looking for evidence against our intended test interpretations" (Haertel, 1999, p. 6). Shepard (1993, p. 429) was concerned that the conceptualization of validity may suggest "the sense that the task is insurmountable" and allow "practitioners to think that a little bit of evidence of whatever type will suffice". Shepard proposed, like Brennan (2001), that a series of appropriate questions be asked and that validity evaluations be organized to respond to questions like: "What does the testing program claim to do?"; "What are the arguments for and against the intended aims of the test"; and, "What does the test do in the system other than what it claims, for good or bad?" (Shepard, 1993, p. 429). Asking questions such as these, will root validity evidence in the specific "local contexts which influence their meaning" (Moss et al., 2006, p. 123), the type of evidence available (Shepard, 1993) and in the description of how the evidence can be used.

Validity studies should happen as regularly as tests are developed and used (O'Neil, Sireci, & Huff, 2003-2004). As validity theory developed, it became clear that multiple sources of evidence were required to support the use of a test for its purposes (Sireci, 2009). Professional associations like AERA, APA and NCME (AERA, APA, & NCME, 1999) called for a variety of forms of validity evidence to support the use of tests for specific purposes. Such evidence should include that inferences obtained from an analysis of <u>test content</u> which is defined as the "themes, wording, item formats, tasks or questions on a test as well as scoring guidelines and administrative procedures for the test" (AERA, APA & NCE, 1999, p. 11). The simplicity of the concept of content validity and its ease of use, continues to be the focus of many teachers in current test practice (Killen, 2003). Such a focus on content relevance and representativeness does not consider student performance and thus ignores "the responsibility that teachers have to interpret the tests in defensible ways" (Killen, 2003, p. 3). "[T]he only way to know whether and assessment really assesses your intended goals is to gather evidence corroborating the test score interpretation" (Herman, Aschbacher, & Winters, 1992, p. 102).

Other sources of validity evidence called for by the *Standards for educational and psychological testing* should include <u>response processes</u>, <u>internal structure</u>, <u>relations to other variables</u> and <u>consequences of testing</u> (AERA, APA & NCE, 1999). In their review of the kinds of sources of validity evidence used in a diverse and large number of educational and psychological tests, Cizek et al. (2008) found disagreement about which sources of validity evidence were considered necessary in supporting test use. These authors also found that the kinds of validity evidence collected were not necessarily consistent with researcher perspectives of current validity theory. For example, while the value, implications and social consequences of a test might not always be considered a crucial aspect of a narrowly defined view of validity (Madaus, 1988), this dimension of the impact of a test is held central in a broader understanding of validity (Messick, 1989a), and accordingly required that suitable evidence be collected. Despite the recognized importance of evidence about the consequences of testing few studies have included this in validity studies of tests (Cizek et al., 2010). Cizek et al. (2010) recommended that validity theory be redefined to clearly differentiate between the validation of test scores (gathering evidence to support score meaning) and the validation of test use (gathering evidence to support the justification of test use).

Despite the three most recent *Standards for educational and psychological testing* (AERA et al., 1985, 1999; APA et al., 1974) stating that it was not tests that are validated, Popham (2006) noted that there persists a popular misconception amongst educators that validity is a property of a test rather than the inferences that are drawn from the test. Neither are test scores validated − it is the interpretations and uses of test scores that are validated (Kane, 2006a, 2009; Sireci, 2009). Validity concerns “the use of a test for a particular purpose” (Sireci, 2009, p. 20). Clearly there is no consensus as to the theoretical meaning of validity or even how the validation of tests should happen (Lissitz, 2009). Gorin (2007, p. 456) said that “despite my best attempts to describe the holy trinity, the unified framework, or argument-based approaches to validity, few students emerge from class with confidence that they could evaluate validity when developing, using, or even selecting tests”. A recent book “*The concept of validity: revisions, new directions, and applications”* brought together the work of a group of eminent ‘validity scholars’ (Lissitz, 2009). In his introduction to the book, Lissitz identified three groups of papers by these scholars: a “mainstream” group whose members are “largely supportive of the literature on validity” (Lissitz, 2009, p. 6), a second group called “relatively divergent” who seem to imply “that the field of research on the concept of validity is taking us in a direction that is just plain wrong” (Lissitz, 2009, p.9), and a third group called “application oriented” because they provide “coherent summaries of actual validations, and perhaps even more important, they are trying to provide advice for those involved in specific application efforts” (Lissitz, 2009, p.11). What this book demonstrated is a range of divergent understanding as to what validation of test and test scores means.

In recent contested work, Lissitz and Samuelsen (2007a) and Borsboom et al. (2009) argued that the validity of a test resides in the test itself and how it was constructed. Lissitz and Samuelsen (2007a, 2007b) questioned the value of a unitary theory which focused on construct validation if such an approach diverts attention from the arguments and procedures associated with content validity. Lissitz and Samuelson (2007a, 2007b) did not imply that the other non-content aspects of validity should be ignored, but rather that it makes little sense to try to meaningfully interpret student performances if content validity is not first secured, because student performance relates to the content of a test. Similarly, Chalhoub-DeVille (2009) acknowledged the importance of content validity when interpreting tests and test scores, but cautioned that on its own, content validity would offer an incomplete and fragmented view of validity.

Irrespective of the different ways that validity has been and is conceptualized, validity evidence is necessary as a basis on which inferences, decisions and arguments about the validation processes are made. There are a number of different techniques that can provide the specific evidence "that is relevant to judging the plausibility that assessment results have a specified set of effects … Once collected that evidence needs to be used in the development of a coherent validity argument" (Linn, 2009, p. 210). The section below describes sources of validity evidence that have been used in the interpretation of test scores. Some of these sources of validity evidence will become part of the conceptual framework (Section 3.5.1) which directs the analyses of the SC Biology examinations, and gives meaning of the SC Biology examinations.

3.2.2 Sources of validity evidence

Collecting validity evidence for use in the validation of a test (examination), is the responsibility of both the test setter and the test user (Nitko & Brookhart, 2007) and the kinds of validity evidence depends on the level of the test, the purpose of the test and the consequences of a test (Cizek et al., 2008). Collection of validity evidence should begin during the development of test items (questions) that comprise the test itself (Downing & Haladyna, 1997), and should embrace the entire examination process (Nitko, 2001, Downing, 2006). Validity theorists have proposed several frameworks for organizing and generating the necessary validity evidence for tests and examinations. Examples of frameworks for investigating validity are those of Messick (1989a), Linn, Baker and Dunbar (1991), and Kane (1992). Downing (2006, p.4) described twelve steps for effective test development and suggested that these steps "provide a convenient framework for collecting and supporting all sources of validity evidence for a testing program". Nitko (2001) and Nitko and Brookhart (2007) combined the validity evidence considered important by Cronbach (1988, 1989), Messick (1989a, 1989b), Linn et al. (1991) and Kane (1992), into a single framework of validity evidence for educational assessments with eight different categories of validation evidence. Prioritizing these categories of validation evidence depends on the interpretations and use of an assessments results (Nitko & Brookhart, 2007). Four categories of validity evidence identified by Nitko and Bookhart (2007) are pertinent to this thesis and are discussed further. These categories are content evidence (content representativeness and content relevance); substantive evidence (types of thinking skills and processes required); internal structure evidence (relationships among the assessment tasks or parts of the assessment); and score reliability evidence (reliability of scores over time, assessors, and content domain). Discussion of these four groups of validity evidence is preceded by a discussion of factors affecting student achievement in examinations (Embretson, 1983) which have been considered important sources of validity evidence but which are not obvious inclusions as content evidence, substantive evidence, internal structure evidence, and score reliability evidence.

Much of the earlier work on assessment validity (e.g., Hambleton, 1984) focused on validity evidence related to content representativeness and relevance. Lissitz and Samuelsen (2007a), Lissitz (2009) and Borsboom et al. (2009) argued for the central importance of content validity evidence for deriving the meaning of students' scores in a test. Others, for example, Kane (2006b, p.135) contended that while "content evidence does not apply directly to test scores, it can support interpretations of scores in terms of expected performance". Sireci (1998, p. 103) argued that one "should not evaluate test scores without verifying [i.e., validating] the quality and appropriateness of the tasks and stimuli from which the scores are derived". At best, a test can only ever be a subset of the content to be taught and learnt by all students. Therefore, the content that is included or excluded from a test must be an explicit contributing element to the specificity of the test and its purposes. While content evidence for validity should not be used on its own, it is, nonetheless, a fundamental requirement of all assessments that should be included with evidence from other aspects of validation (Sireci, 1998).

Abedi (2002, 2006) described other factors about tests which are not directly related to the content of the test but influence the content of a test. Of these factors, Abedi (2006) considered linguistic complexity of the test to be the most influential because it increased the difficulty of a test especially for test-takers with a home language other than the test language (Abedi, 2006; Martiniello, 2008, 2009). Unnecessary linguistic complexity is associated with construct-irrelevant threats to validity for all students, especially for English Language Learners (ELLs) (Abedi, 2006; Martiniello, 2008, 2009). Non-linguistic components such as diagrams, tables and graphs have been shown to be important in determining the difficulty level of mathematics questions, and therefore the validity, of mathematics tests (Martiniello, 2008, 2009). Like mathematics, science tests comprise both linguistic and non-linguistic components individually and sometimes in combination, and non-linguistic communication is considered to be extremely important is science (NRC, 1997, 1999). Therefore, collecting validity evidence about linguistic complexity, such as the structures of questions and their expected answers and whether questions and answers make use of non-linguistic components, is essential in constructing a validity argument. In addition, as general characteristics such as, for example, the length of a test and the number of items in a test, also influenced the difficulty of a test (Britton et al., 1996a; Abedi, 2006), validity evidence should include exploring these elements too.

Validity arguments should not be confined to the test itself, as has been discussed thus far, but also include the score (mark) interpretations and uses (Frisbie, 2005), that is, understanding student performance. The most important emphasis in the test scoring step is accuracy (Downing, 2006) because scoring errors in final test scores reduce validity evidence (Haladyna & Downing, 2004). Britton et al. (1996a) found that differences in the structure of examinations, such as the length of questions and expected answers, and differences in emphasis of different topics, impacted students' achievement. Summary test statistics, for example raw scores, means and standard deviations are important validity evidence and must be evaluated and documented (Downing, 2006). Any anomalies identified by the validation processes associated with a test should be investigated and resolved prior to reporting test scores (Downing, 2006). This is particularly important in the absence of more sophisticated statistical methods of determining the equivalence between tests (Downing, 2006). Therefore, because of their effects, factors such as these anomalies and their resolutions need be included when collecting validity evidence.

"Measuring simple factual knowledge across the content domain is not a substitute for measuring comprehension of concepts or skill in using content knowledge or skill in using content knowledge to solve novel problems" (Linn, 2006, p. 30). The *Standards for educational and psychological testing* (AERA, APA, & NCME, 1999) require evidence to support (or not support) interpretations of test scores with respect to cognitive processes. Test validity thus depends in part upon whether the test simulates real-world "tests" of ability in the domain being tested (Wiggins, 1990). How we define "ability" can be problematic and confusing, especially if comparisons are to be made within and between tests, and there is no common understanding of what ability implies. Understanding the cognitive demand of an assessment helps test-users make defensible inferences about test-takers, cognitive abilities from their responses to different questions (Pellegrino, Chudowsky & Glaser, 2001). One way of achieving a common description of abilities, as is required when generating validity evidence, is by using an organizing framework or a taxonomy which enables abilities to be categorized, in terms of, for example, the cognitive challenge of a question according to particular criteria or the performance expectations of a question.

One of the first documented ways of classifying cognitive skills is the original taxonomy of Bloom and his colleagues (i.e., Bloom et al., 1956) called the Bloom's Taxonomy of Educational Objectives, or the BTEO. What this taxonomy provides is a standard vocabulary which can be used to describe what any educational objective, and therefore any question, intends to measure in the cognitive domain. While the Bloom et al. (1956) taxonomy was never intended to offer more than a framework for use by different disciplines or fields according to their specific needs, it has became one of the best-known ways of understanding student learning (Anderson et al., 2001). This taxonomy categorized the cognitive domain as either Knowledge (least complex), or Comprehension, Application, Analysis, Synthesis or Evaluation (most complex). In spite of being revised (Anderson et al., 2001), the original BTEO remains popular. For example, in a paper recently published in *Science* (Zheng, Lawhorn, Lumley & Freeman, 2008), the authors used the original BTEO as a tool to evaluate the critical thinking required by some university placement and medical admissions tests in the USA. Various validity studies such as, alignment analyses, have used different methods (i.e., criteria or taxonomies), or combinations of criteria and taxonomies, to evaluate the level of cognitive demand of assessment questions and the standards against which the questions are being compared (e.g., Webb, 1999; Porter, 2002). The choice of measure is dependent on the purpose for the type of validity evidence required. Recognition of the importance of cognitive development and utilization in education, especially for the twenty-first century (Oakland, 1995), has led to the development of a number of different ways in which students' cognitive growth and development can be understood and measured. However, there is a worldwide lack of consistency in how distinct studies measure and understand the cognitive demand of assessments (Schraw & Robinson, 2011).

Currently in South Africa, the original BTEO is used to set the SC Biology examinations in accordance with the guideline documents (DoE, 2001a, 2002a). In South Africa, educators at all levels continue to experience problems using either the original or the revised BTEO, because of elements of subjectivity associated with their use. For example, the national (state) policy since 2001 (DoE, 2001a) has designated the BTEO Application as a higher order skill (DoE, 2001a), yet one of the independent examining bodies, the IEB, treats BTEO Application as a lower order skill (IEB, 2007a). This contrast is a critical difference when comparing the DoE and the IEB SC Biology examinations. While there are a number of different ways in which researchers have classified cognitive demand, there is little consistency between the results these ways yield when used (Schraw & Robinson, 2011) (see Chapter 4). For this reason the author of this thesis developed an objective instrument to assess cognitive demand for use in this study. The development of this taxonomy, called the Performance Expectations Taxonomy (PET), is described in detail in Chapter 4, where the PET is compared with other methods used internationally for assessing cognitive demand.

A number of USA validity studies have included evidence of the cognitive demand of tasks in alignment studies between different components of the curriculum. Alignment has direct effects on the degree to which valid and meaningful inferences can be made from student assessment data (Long & Benson, 1998; Rothman, 2003), especially when a standards-based curriculum operates. Alignment as a source of validity evidence has developed in recent years as a result of the introduction and the use of content standards and performance standards, together with legislation in the USA which requires specifically that assessments be aligned with the standards (Webb, 2007). Alignment will be discussed separately in detail below because particular methods used in alignment analyses to generate validity evidence are also used in the conceptualization of the framework developed to explicate standards from the SC Biology examinations (Section 3.5.1) and in operationalization of the conceptual framework (Chapters 4 and 5) which guides this study.

3.3 **Alignment**

It has been argued that alignment has existed in one way or another almost since the beginning of the use of formal tests to aid decision making" (Impara, 2001, p. 3). There are a number of different definitions of alignment in the literature. Alignment has been described in a fairly broad way as "the degree to which various policy instruments available to the system such as standards, textbooks, and assessments accord with each other and with school practice" (Schmidt et al., 2005, p. 527) to achieve the "desired goals of stakeholders" (Case & Zucker, 2005a, p. 1), or the "single goal, educating students to reach high academic standards" (Hansche, 1998, p. 21). Newmann, Smith, Allensworth and Bryk (2001) argued that improved student achievement would be more likely if all facets of a school's instructional program were aligned with external policies on a common framework such as standards. Porter (2006, p. 147) gave the simplest explanation when he described alignment as the extent to which "content is the same" between the elements of the education system being compared. Alignment therefore measures the overlap or similarity between the elements within an education system, and between education systems, with respect to content.

While the conceptualization of alignment arose from a systemic focus, many of the explanations of alignment have focused particularly on the role of assessments and USA standards (e.g., Bhola, Impara & Buckendahl, 2003;Webb, 1997a, 1997b). The need for alignment studies arose in the USA to monitor the influence or effects of standards on the other parts of the education system. Government legislation in the USA, namely, 'The No Child Left Behind (NCLB) Act' of 2001, has resulted in efforts focused on developing mechanisms for determining alignment, specifically between standards, assessment and instruction (Herman & Webb, 2007). In the USA, alignment models have been developed in response to the NCLB Act which stipulates that individual states show evidence that their assessments are in line with their state standards. Organizations faced with the task of monitoring the required alignment then need to determine, first, how to decide if standards are equivalent between states, and, second, how to discern the equivalence of the alignment measures as generated by the different alignment models.

Alignment between standards and an assessment has also been recognized as increasing an assessment's reliability (Case & Zucker, 2005b) and evidence of an assessment's validity (Case, Jorgensen & Zucker, 2004; Case & Zucker, 2005a; Impara, 2001; Resnick, Rothman, Slattery & Vranek, 2003; Webb, 1997a), as this alignment indicates how well an assessment is measuring what it is intended to measure – students' performance in relation to the standards. While alignment judgments and scores are used to validate assessments with respect to particular standards, the value of these measures depends on how the concept is operationalized, that is, the particular type of methodology that is used to perform the alignment (Rothman, 2003). There are different approaches to measuring alignment, some of which have been reviewed by, for example, Bhola et al. (2003), the Council of Chief State School Officers (CCSSO) (2002), La Marca, Redfield, Winter, Bailey and Despriet (2000), and Rothman (2003). To be perfectly aligned, both the standards and assessments should focus consistently on the same criteria, such as, for example, on particular content or skills. The consistency of alignment can then be measured by evaluating or measuring how many or how few of the criteria are shared between a point of reference, for example, standards, and an assessment. The degree of alignment can vary and will contribute to evidence of validity.

Most USA alignment studies involve comparing each assessment to clearly articulated standards, rather than comparing assessments directly with each other and "do not address the quality of the standards, the assessments or the individual test items" (Beck, 2007, p. 134). The alignment methodology developed by Porter (2002) differs from that of, for example, Webb (1997a), because Porter compared the standards and assessments separately, quantitatively and visually using a common framework, a content-by-cognitive demand matrix, rather than directly against each another.

Most alignment methods have included comparing types and the weighting of content and substantive (cognitive demand) validity evidence, as recommended by, for example, Nitko (2001) and Nitko and Brookhart (2007). Substantive validity evidence (i.e., depth of knowledge or cognitive challenge) can be collected using a number of different measures, sometimes called taxonomies. Most often it is the level of cognitive demand that is used as an indication of the intellectual challenge of an assessment. It is thus important that ideally the most suitable way of categorizing this variable be determined or that, at least, a robustly defensible categorization is sought and adopted. The development of a tool to measure the cognitive demand of SC Biology examination questions is the subject of Chapter 4.

As alignment methodologies have conceptually and practically concerned comparisons between different part of education systems, including assessments and standards, they offer practical ways in which to compare SC Biology examinations through the lens of standards. Therefore, alignment has been incorporated into the conceptual framework of this study (Section 3.5.1) and appropriate alignment methodologies have been incorporated into the operationalization of this framework (Chapter 5).

3.4 Equating of assessments

Score or mark users, such as admissions staff at universities and employers, need to be able to compare the marks of examinees who took different 'forms' of a test (or examination)[96] or different tests at different times, in order to make decisions about people. There is a need in studies of the comparability of assessments for judgmental processes which would result in consensus as to whether the performance in different sets of tasks or tests were of equivalent rigour or not (Linn, 1995). Ideally comparability between tests should start with test construction (Mislevy, 1992). Careful review and analyses of the individual items that comprise each test can produce tests that are very similar to one another in terms of content and difficulty, but such reviews do not necessarily guarantee equivalence (Wendler & Walker, 2006) and do not take into account differences in the relative abilities of the different populations of test takers.

There are a number of different ways in which validity evidence about equivalence can be collected, such as, for example, through the processes of *equating*, *linking* and *scaling* (Brennan, 2006a). Validity evidence about equivalence can then be used to compare tests, comparable scores and more accurate estimates of ability. A review of the various statistical methods used to determine equivalence between different examinations is beyond the scope of this thesis. Different methods vary conceptually, and in terms of their complexity and the procedures which they use. Due to space constraints, a detailed review of all the procedures concerned with establishing the equivalence, or lack of equivalence, between assessments will not be covered here. Interested readers can, for example, consult Brennan (2006b); Kolen (2006); Holland and Dorans (2006); Jones, Smith and Talley (2006); Kim, Walker and McHale (2009); and Livingston (2004). Brief

[96] In a South African SC context, different forms of a test could refer to: one examination (one examining authority) with one or more choices between questions; and/or different examinations (more than one examining body) written in the same examination cycle: and/or different examinations in different years. The existence of different forms of tests challenge decisions about equivalency and comparability (Linn, 1995).

explanations of equating, linking and scaling are provided below because of their philosophical impact on various forms of equivalence that have been assumed by the users of SC examination scores in South Africa. More discussion is devoted to scaling because the South African SC examinations were norm-referenced (NEPI, 1992b).

Equating refers to statistical methods that allow the scores from one test to be converted to the scores on another test, and which compensate for small differences arising from small variations in the difficulty of the tests resulting in comparable and interchangeable scores for both tests (Jones et al., 2006; Wendler & Walker, 2006). Brennan (2001) indicated that sometimes there is a mismatch between the theory of equating and how it is practiced. They suggested that to qualify as elements for equating procedures, the tests concerned should be constructed according to the same content, the same statistical procedures and be administered under the same conditions (Brennan, 2001). If the tests do not satisfy these conditions, then the term *linking* should be used (Brennan, 2001) to convey a connectedness that does not satisfy equivalence criteria.

Linking refers to the equating of multiple generations of test forms across time (Jones et al., 2006) or transforming the scores from one test to the scores of another test (Holland & Dorans, 2006). For example, linking is necessary when a new test is set for each test administration or test cycle, and each test is purposefully designed to be different. Therefore, different examinee populations and different ability groups must be controlled for. Brennan (2001, 2006a) indicated the role of linking in public and institutional comparisons of different tests with similar purposes, especially those with identical or similar titles. Linking is problematic because it seeks to relate "scores for different measures and/or different conditions of measurement (Brennan, 2001) which means that scores between tests cannot be used interchangeably (Kolen & Brennan, 2004). Linking is practiced in different ways and ways of linking "are evolving at a rapid rate" (Kolen & Brennan, 2004, p. 424). One way of linking is, for example, the use of a set of common questions, called anchor questions in both tests. Statistical procedures then remove the non-equivalence between the tests. This method of linking makes additional assumptions (Holland & Dorans, 2006) which will not be discussed in detail here. The new conditions which might arise out of new test content could necessitate new standard-setting procedures (i.e., setting new cut-scores) rather than an equating (Jones et al., 2006). When tests include very different content, linking procedures will not suffice to ensure equivalence of the tests (Kolen & Brennan, 2001).

The role of s*caling* is to create scores that can be used and interpreted by test users (Jones et al., 2006) or to give meaning to score scales (Kolen & Brennan, 2006; Kolen, 2006). If scales are used to confer a practical meaning or understanding of test scores, the scales need to be reported in ways that have meaning for teachers and other test score users. In order to do this, test scores "derive meaning by being linked to some [relevant] reference group or some specific criterion" (Wendler & Walker, 2006, p. 463). A norm-referenced scale[97] attempts to give meaning to scores by "comparing a test-taker's performance to that of persons in the comparison or norm group" (Linn, 2006, pp. 35, 36). When a criterion- referenced scale[98] is used, the interpretation of a student's score is made independent of the performance of anyone else, "and would be associated with a criterion or criteria for judging success on the test that would have been specified a priori" (Cizek & Bunch, 2007, p. 334). Popham (2009, p. 102) described criterion-referenced measurement as "an approach to testing based on how well the test-taker has mastered a well-defined criterion behaviour, such as a cognitive skill or body of knowledge". A criterion-referenced scale could refer to performance categories such as, for example, advanced, proficient, basic, or simply pass and fail (Linn, 2006), or might reflect performance relative to established standards (Kolen, 2006) in which case the term standards-referencing is used (Cizek & Bunch, 2007). Criterion-referenced scores are valued because interpretations made about a student's performance in a test are based on some criteria which can give feedback to a student as to how he/she can improve (Cizek & Bunch, 2007), provided the appropriate criteria are identified and built into the test. Score scales can change their meaning over time due to shifting test populations. Changes made to test content specifications, and small changes made to content which accumulate over time (i.e., test content drift), may necessitate rescaling (Wendler & Walker, 2006). The development and setting of score scales is part of the standard-setting process. Brennan (2001) argued that explanations of score scales, especially the assumptions involved in the method by which scale scores were derived, are essential in the test validation process.

In the literature, norm-referencing and criterion-referencing are used as being distinct from each other, yet depending on how they are described conceptually or interpreted, each can include elements of the other. For example, Tognolini and Stanley (2007, p. 130) wrote that:

> [o]ne of the main advantages of norm[-]referencing is that the marks, grades or awards are interpreted in the same way from situation to situation (year-to-year; subject-to-subject) … [t]his means for example, that each year 20% of the students in a subject will receive a distinction and 30% will fail.

This particular advantage presupposes a static state of performances, and a somewhat inane equivalence. It would also require that exactly the same norm group was used every year and for

97 Haertel (2006) referred to norm-referencing as relative because decisions about scores depended on the relative ordering of students.

98 Haertel (2006) referred to criterion-referencing as absolute because decisions about scores are made some absolute criterion.

every subject and that no consideration be given to the content of each examination. Stanley and Tognolini (2008, p. 13) themselves noted that for the examinations in the subjects that they studied in both the UK and parts of Australia, the descriptors for high achievement in the examinations between subjects had "semantic similarity" and it was "easier to interpret similarity within the subject discipline than across the subject disciplines". Despite defining standards-referencing as referencing achievement to pre-determined standards of performance, Stanley and Tognolini (2008), considered standards-referencing to be different from criterion-referencing.

Tognolini and Stanley (2007, p. 142) stated that a weakness of norm-referenced tests is that when weak students ask how they can improve "they are told that they can only improve by beating other students in the class". It could be argued that this response to a student assumes that the student would only be considered improved if they changed their rank in the group of test-takers. An experienced teacher could equally have said to a weak student that they can improve their mark by doing x, y and z. Having a norm-referenced test does not preclude a teacher from analyzing student performances which reveal students' strengths and weaknesses.

It is difficult to imagine how, when one purpose of a test is some form of selection of students for a particular purpose, the ranking function of norm-referencing can be avoided, even if the test claims to be a criterion-referenced test. Consequently, "[n]ormative information needs to be made part of the process for judges to anchor their absolute [criterion] judgements with an understanding of current levels of performance of students and likely consequences" (Linn, 2003, p. 14). Cizek and Bunch (2007, p. 11) capture this sentiment more simply by writing that "criterion-referenced judgements cannot be made without at least implicit consideration of normative information".

Therefore, scaling, in particular whether the choice between a norm-referenced scale or a criterion-referenced scale is used, depends on how students' marks in a test are to be reported. Student performance in a test can be reported (ranked) by the score each student achieved with respect to other students (norm-referenced), and/or can be reported relative to some domain or standard (criterion-referenced). In a standards-based curriculum when a student's score is related to a particular standard (content and/or reference standards) it is called standards-referenced. Equivalence between tests and examinations should in principle be potentially easier to establish in a standard-based curriculum given the common and consistent framework provided by the standards to a curriculum (see Section 3.1.5).

3.5 A conceptual framework to explicate standards from South African Senior Certificate Biology examinations

3.5.1 A description of the conceptual framework

The conceptual framework summarized in Figure 3.5 involves all of the concepts previously discussed in this chapter, and is used to elicit the standards of the SC Biology examinations from question papers and candidates' answer scripts. How these concepts relate to each another, and to this study in particular, forms the focus of the discussion below.

Validity is about giving recognition of the meaningfulness of the student scores in assessments like examinations (Cizek 2001). Standards exhibit a society's expectations of students and student scores are indicators of whether students have or have not met the standards in particular assessments. That is, standards─both content standards and performance standards─imply meaningfulness of student performances. In an education system that is standards-based, explicit standards would indicate what it is that students must know and be able to do in order to be determined competent from an examination. If the standards are not explicit then they need to be explicated in order to extract or infer what competency in student performances means. In essence, validity is about evidence-based trustworthiness and the kind of evidence depends on the context. Validity evidence must be assembled to determine whether the examination process does in fact determine the competencies (or standards) that it claims to assess, and does it classify students accordingly.

The SC Biology examinations in South Africa, are a collection of tasks or questions used to rate or judge a candidate's performance, or mastery, at a particular level. The mark or grade awarded on an examination script on the basis of a candidate's performance represents his/her mastery in identified and explicated collective competencies.[99] Mastery is expressed in terms of performance standards (Figure 3.5[*i*]), which only derive their specific 'meaning' from their relationship to the content standards (Figure 3.5[*ii*]). In this study content standards are implicit in the examination question papers (Figure 3.5[*iii*]). The rigour of the content standards depends of the cut-scores for levels of student mastery, that is, the performance standards (Figure 3.5[*ii*]). The way in which content standards and performance standards are conceptualized in this study (see Section 3.5.3) draws on the way they are both practiced in the USA (Figure 3.5[i*v*]) and how performance standards are practiced in the UK-Australia-Netherlands (Figure 3.5[*v*]). Content standards are generated by analyses of examination question papers (Figure 3.5[*vi*]) and the performance

[99] "It is necessary to use performances to infer competence, but competence transcends a particular performance. Moreover, performance and competence are not in a perfect one-to-one relationship; instead, the relationship involves indeterminacy and imperfection" (Andrich, 2002, p. 38).

standards are generated by analysis of student answer scripts (Figure 3.5[*vii*]), together with the content standards of the appropriate question papers (Figure 3.5[*viii*]).

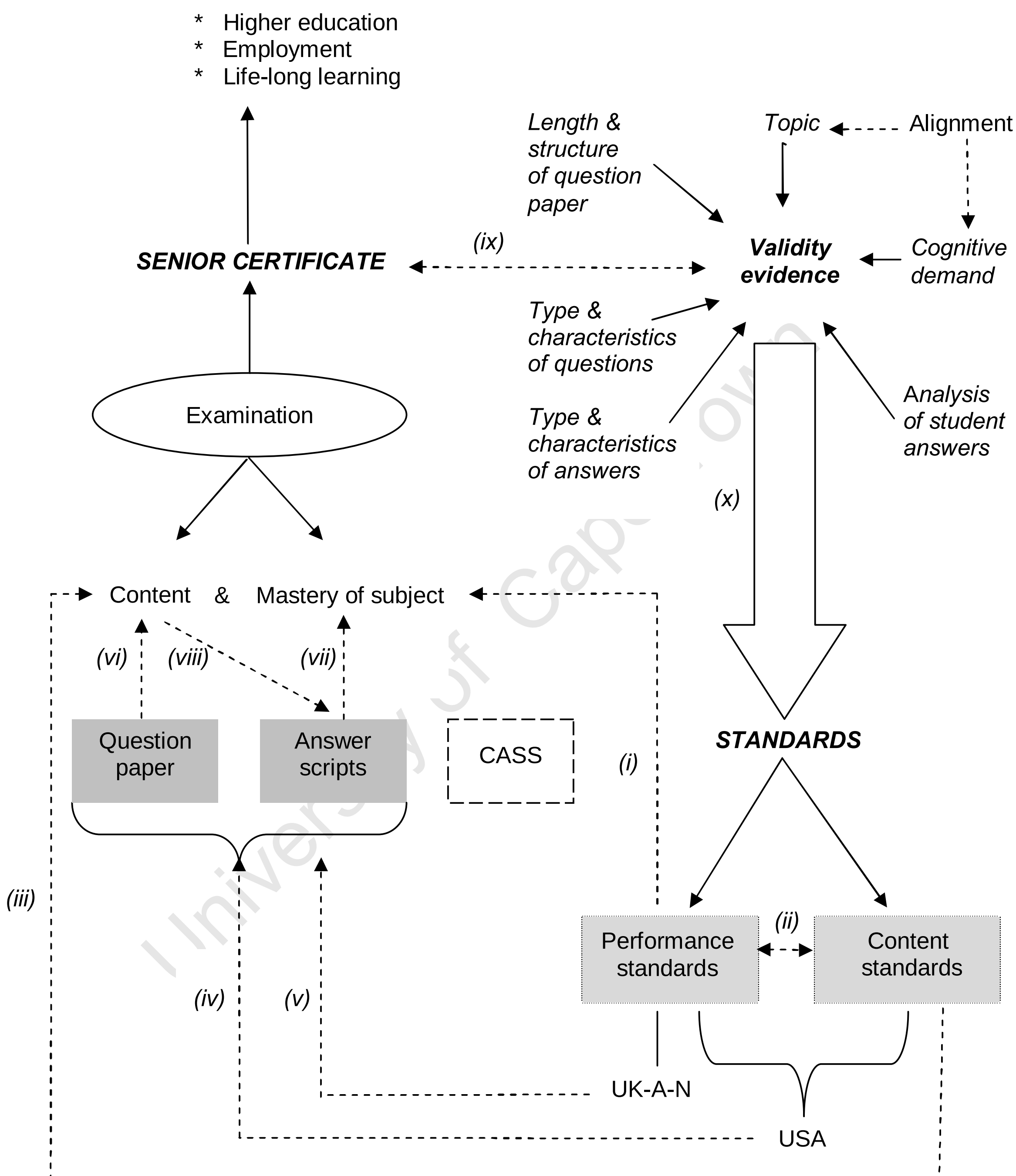


Figure 3.5 An overview of a conceptual framework for understanding SC Biology standards by examination analyses. Broken lines represent relationships operationalized in Chapter 5. UK-A-N refers to the UK-Australia-Netherlands grouping described in the text.

Validation of a South African SC Biology examination (Figure 3.5[*ix*]), should require that different levels of mastery reflect performance (performance standards) (Figure 3.5[i]) in a particular set of knowledge and skills (content standards) (Figure 3.5[iii]) considered to be important by the society that recognizes them or uses them. Therefore, in the absence of explicit standards, as in South Africa, validity evidence may still be collected *post hoc* but not used to determine if a set of standards have been met (or not met) because there are no standards against which to compare the evidence. Instead, the validity evidence may be used to silhouette the implicit standards for a particular examination because the validity evidence will reflect the knowledge and skills that educational decision-makers assumed would meet South African society's expectations, the standards (Figure 3.6[x]). That is, validity evidence gives meaning to the standards of each SC Biology examination.

The kinds of validity evidence necessary for validation would include information that can be generated by analyses of both examination question papers and of the marks obtained by students through analyses of student answers scripts. In this thesis, Brennan's model of Socratic validation (Brennan, 2001), which has been previously described, is used. That is, validity evidence is sought to address specific questions about the way test scores are used to distinguish, or not distinguish, between students in South African SC Biology examinations. Here, addressing specific questions refers to interrogating whatever sources of validity evidence (see Section 3.2.2 above) are available for analysis, in the light of the research question and sub-questions posed in Chapter 1. Analyses of examination question papers should yield validity evidence about the length and structure of a question paper, the types and characteristics of question within the question paper, the type and characteristics of the answers students are required to give (structural aspects of the examination question paper). The topics and cognitive demand required by the questions reflect the implicit content standards. Analyses of student answer scripts in the light of content standards generated for a particular question paper will yield explicit performance standards. The validity evidence is organized in three strands that run through the organization of Chapters 4 to 6. These strands are called: the structural aspects of question papers; the content standards; and the performance standards. Chapter 5 describes the methodology for analyzing these three strands of evidence from the SC Biology examinations and in Chapter 6 the SC examinations, are compared with respect to these three strands of evidence.

As no two SC examinations are exactly the same, the standards generated in this way may vary between examinations. But because the standards are generated by the same validity evidence framework, the standards can be analyzed, described and compared within and between years. Standards generated via validity evidence satisfy the seven aspects considered to define education standards in Section 3.1.1 in the following ways. Therefore, for any specific SC examination the validity evidence (standards) described (**measured**),[100] as an indication of the **quality**, the topics and performance expectations (**criteria**) assessed (**practiced**) in the SC examinations, in the light of curriculum documents and examination question papers developed by subject specialists appointed by the South African Government (**recognized authority**) and of corresponding student **achievement** in these examinations (answer scripts). Validity then becomes a lens to view SC examinations and by which the explicated standards can be described, generated and compared. Comparisons of SC examinations may or may not be linked to changed policies (**revised**).

3.5.2 The significance of the conceptual framework for this study

Various forms of validity evidence about examinations should be assembled to support their use for specific purpose(s) and for any interpretations which might result from corresponding students' performances. Candidates who were successful in the South African SC examinations were certified (the purpose) to be differentially competent (performance standards) in particular content knowledge and skills (content standards), as defined in curricula documents, and successful students were considered to be qualified to enter the workplace and/or to continue with tertiary education (the interpretation).

This study seeks to investigate changes in the SC Biology examinations by the analyses of SC Biology question papers and supplemented in some examination years by candidates answer scripts through the lens of standards. In the absence of a standards-based curriculum in South Africa, the conceptual framework developed in this study developed a way to extract standards from the SC Biology examinations in the country. By articulating explicit standards, using consistent and coherent procedures, using validity evidence as a proxy for unrecorded standards, the inferred content standards' and performance standards can be described, analyzed and measured in any year and compared between years. The conceptual framework is operationalized in Chapters 4 and 5, and the content standards and performance standards of selected SC Biology examinations are described, and compared in Chapters 6 and 7. Standards generated through this conceptual framework also provide an explicit benchmark against which NSC Biology examinations in the post-2007 curriculum can be compared to similar examinations of previous years.[101]

[100] Terms in bold represent the aspects of standards developed earlier in this paper (Section 3.1.1) that can be derived from question papers and answer scripts.

[101] The framework developed in this study has been used to compare the NSC Life Sciences examinations for 2008 to 2010, with the analyses of SC Biology examinations completed in this study (Crowe, in preparation). The analyses have been excluded from this thesis because the context of the Life Sciences curriculum is different to that of Biology.

Equivalence between SC Biology examinations within and between years is conventionally assumed to be able to be achievable in post-examination processes of Umalusi, the current assessment standards authority for school-leaving examinations, and its various predecessors. These processes use statistical procedures which adjust the total scores of a subject in the light of a moving norm composed from the previous three to five years raw score history of the same examinations. The marks of students who do not write English as a first language have their raw scores statistically adjusted upwards to compensate for their potential language disadvantage (Fatti, 2006). Such statistical adjustments do not specifically take into account the differences in content (in this study, topics and performance expectations) which was examined each year or the impact of the structure of examination question papers on student performance. Other assumptions made about student performance in SC Biology examinations involve the assumed equivalence of the some HG performance standards and some SG performance standards. This thesis does not test the validity of each of the SC Biology examinations, instead it examines the assumed validity of inferences made about these examinations in term of standards. Articulating the content and performance standards through the conceptual framework developed in this study allows the notion of equivalence between the examinations within and between years to be investigated.

The conceptual framework recognizes three strands of validity evidence that are important in describing the standards of SC Biology examinations, namely, the structural aspects of question papers, the content standards (topics and cognitive demand), and the performance standards. Variables reflecting the structure of question papers, the structure of questions, the structure of answers, and the topics examined are all relatively objective to measure, and how these variables are measured is described in Chapter 5. Methods of assessing the cognitive demand required by questions within an examination is highly subjective. This study did not use cognitive laboratories to determine the cognitive demands of different questions. Instead, analyses of the expected performances of various questions were used as indicators of cognitive demand. Therefore, an objective instrument, called the Performance Expectations Taxonomy (PET), is developed in Chapter 4. Once developed the PET becomes a part of the conceptual framework described in this chapter.

3.5.3 Working definitions of standards used in the framework and from here on in this thesis

- Standards are the knowledge and skills (*content standards*) that we as a society expect of our students and the modes by which their proficiency therein is recognized (*performance standards*).
- The *content standards* refer specifically to what students should know (topics and procedures) and be able to do with what they know (cognitive demand).[102]
- The *performance standards* span a performance continuum of different performance levels, each of which is defined by differential performance in the content standards.
- Different performance levels are identified by *performance level labels* (symbols), which in South Africa are denoted by the *same cut-score boundaries* for *each examination, each year*. It is unknown if descriptions of the different performance levels, called *performance level descriptions,* and therefore the performance standards, would be equivalent between examinations and between years. In fact, anecdotal evidence would suggest that such equivalence is unlikely in the South African context.

3.6 **Chapter summary**

The first part of this chapter 'unpacks' the conceptual meaning of educational standards as they are understood in international literature. Educational standards are viewed as a social construct, which is recognized by society in by two inter-related groups of attributes. The first group of attributes concerns the establishment and maintenance of standards over time, and the other group is concerned with how standards are operationalized or practiced in education systems. The second part of the chapter describes how standards, in particular content and performance standards, are conceptualized and how they operate within the education systems of some selected countries around the world. These countries are, the USA, the UK-Australia-Netherlands, and South Africa. The benefits of a standards-based curriculum are that, because the standards should bring a coherent and consistent message of what it is important for teaching, learning and assessment, student performance is easier to understand. Standards also provide a framework or language by which comparisons can be made between corresponding examinations taken in different years. As this thesis is a comparative study of the SC Biology examinations, each offered at HG or SG levels, over a period of fourteen years how the work views equating examinations is discussed. The notion of alignment is also invoked, because it has been a criterion used to make comparisons between assessments in the USA.

---

[102] Called performance expectations from Chapter 5 onwards.

Education in South Africa does not follow a standards-based curriculum and it is therefore necessary to construct a conceptual framework by which standards could be explicated from South African SC Biology examinations─the focus of this study. The conceptual framework developed here addresses in part, research sub-question 1 of this thesis (Chapter 1). The operationalization of this conceptual framework in Chapters 4 and 5 completes research sub-question 1. This conceptual framework used aspects of validity of the SC Biology examinations as a proxy for standards, and validity evidence is thus used to generate three strands of validity evidence necessary to impute standards here. The first two strands are the content standards generated from the examination question papers and the performance standards from the examination question papers and candidates' answer scripts. The third strand concerns the structural aspects of question papers known to influence student performance, and therefore the standards of an examination. The significance of the conceptual framework developed here, and the working definitions provided for content standards and performance standards, is that they provide a means for explicating the standards implicit in the SC Biology examinations and may be used comparatively.

CHAPTER 4

DEVELOPMENT OF AN INSTRUMENT TO DETERMINE COGNITIVE DEMAND

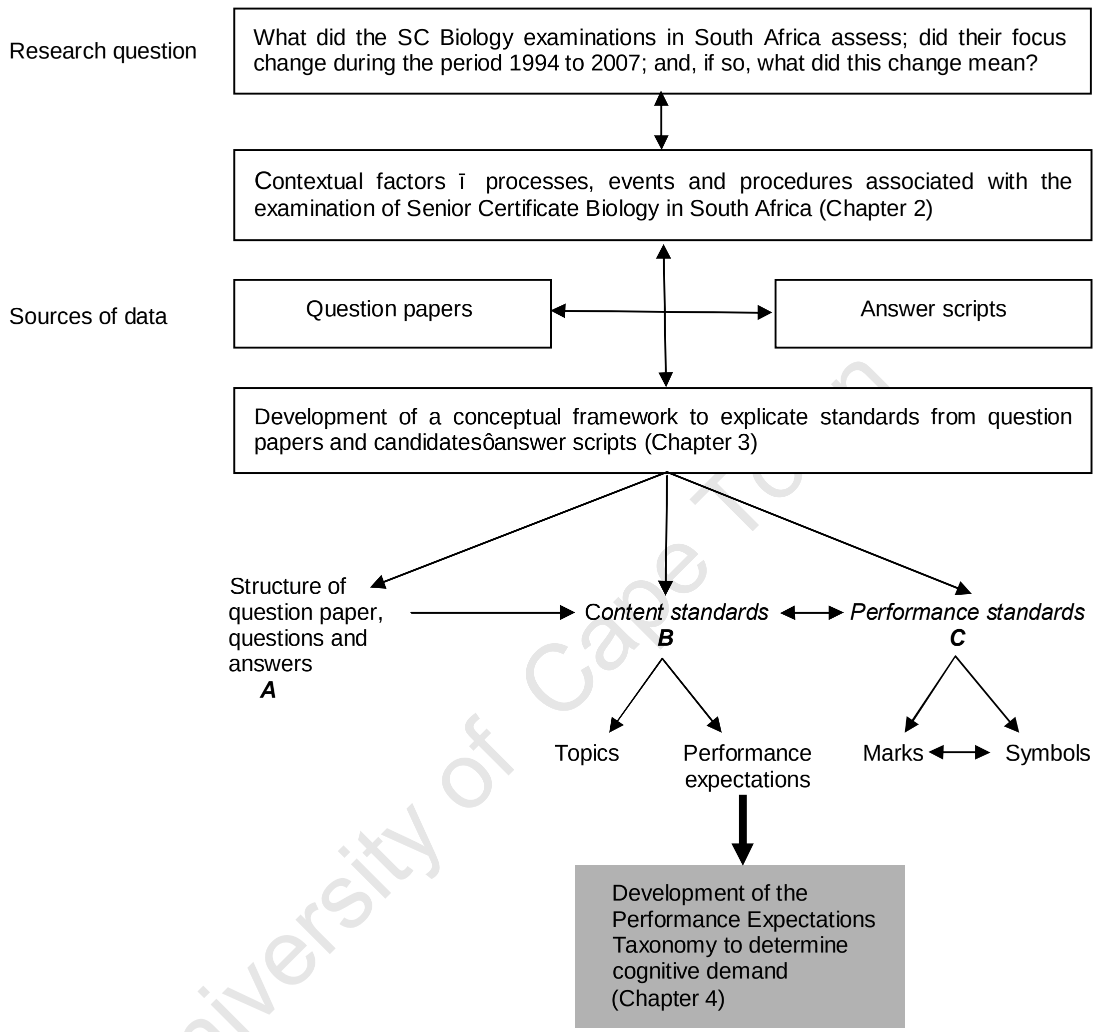


Figure 4.1 The relationship between Chapter 4 and the previous chapters.

The purpose or use of an assessment, like the SC examinations, influences the way it is designed and how its validity and reliability are evaluated (Shepard, 2003). Validity evidence about assessments and the assessment processes, together with relevant theory, support or do not support the interpretation of student scores therein for specific purposes (AERA et al. 1999). That is, validity confers meaning to student performances in tests (Cizek et al., 2008). In Chapter 3, the conceptual framework developed to guide this study identified three strands of evidence as important in understanding student performance in the SC Biology examinations and hence for describing, understanding and comparing student performance therein. These three strands are the structural aspects of question papers (Figure 4.1[*A*]); the content standards (i.e., topics and performance expectations) (Figure 4.1[*B*]); and the performance standards (Figure 4.1[*C*]). The structures of a question paper as a whole, of the questions which are asked in the question paper and the expected student answers to these questions, as well as the topics covered by a question, can all be subjectively determined. The remaining component of content standards, recognized as an important source of evidence in this study, is performance expectations, or cognitive demand.

In the last sixty years, sophisticated and complex models of cognitive processing and complex measurement tools have emerged from a "cross-fertilization" between cognitive psychology and psychometrics (Schraw & Robinson, 2011, p. 12). Despite this background, it is unlikely that there is a single "all-inclusive, all-purpose tool" to measure cognitive demand (Furst, 1981, p. 451) and cognitive demand remains difficult to measure, and to measure consistently in practice (Schraw & Robinson, 2011). A survey of the literature revealed a number of different tools, some of which are called taxonomies,[103] used to classify or group assessment (examination) questions by explicated characteristics which describe different levels of cognitive demand. Taxonomies of cognitive demand vary depending on the type of study in which they are being used, and each taxonomy tells its own story about cognitive demand, depending on its theoretical foundations, and thus yields different classifications of cognitive demand. This variety means that taxonomies may not be used interchangeably within and between studies, and yet they sometimes are in the literature. This study required a principled basis on which to make inferences about the cognitive demand of examinations. This chapter explains the reasons why it was necessary design and develop the Performance Expectations Taxonomy (PET) to determine cognitive demand in this study, rather than to use an existing taxonomy.

This first part of this chapter examines the policy requirements with respect to cognitive demand in South African SC Biology examinations during the period 1994 to 2007 as stated in the CBS (Section 4.1). The second part of this chapter describes some different taxonomies used to determine cognitive demand which have been used in the literature, and explains why they were unsuitable for use in this study (Section 4.2). Bloom's Taxonomy of Education Objectives (BTEO)[104] (Bloom et al., 1956) is discussed in more detail than the other

103 In this context a taxonomy is a descriptive framework, or "creative narrative" (Shulman, 2002, p. 44) which guides the classification of different kinds of cognitive demand "into groups so that each group is as different as possible from all other groups, but each group is as homogenous as possible" (Moseley, Baumfield, Elliott, Gregson, Higgins, Miller & Newton, 2005, p. 37). The author is cognisant of criticisms that some taxonomies of cognitive demand should be called a classificatory system rather than a taxonomy (Krathwohl, Bloom & Masia, 1964). For consistency in this thesis, taxonomy includes all the classification systems discussed.

104 Also referred to as Bloom's Taxonomy.

taxonomies of cognitive demand because it was used for setting the SC Biology examinations. This section includes a discussion of how BTEO has been misrepresented and misused in current literature, especially with respect to the categorization of higher order skills. The third part of this chapter details the development of the PET designed to determine cognitive demand in this study and how the PET was validated (Section 4.3). This part also describes how the PET relates conceptually to the BTEO and two other interpretations of the BTEO (i.e., Crowe, Dirks & Wenderoth, 2008; Zohar, Schwartzer & Tamir, 1998) that have been modified for use in studies of Biology curricula (Section 4.3.3.1).[105] Section 4.3.3.2 calibrates the PET and the BTEO in two ways: first, by cross-referencing the performances expectations of the PET between the two taxonomies; second, by comparing classifications of SC Biology examination questions made by the examiners using the modified CBS version of BTEO with classifications of the same questions made using the PET. The fourth part of this chapter relates the general objectives and principles given in the CBS to the cognitive demand categories of PET (Section 4.4). Selected examples of SC Biology examination questions classified using the PET, together with the rationale for the particular classification of each, are given in Section 4.5 to illustrate the how PET works.

This chapter does not attempt to critique the current evolving conceptions of learning (e.g., Schraw & Robinson, 2011) as these topics are not part of this thesis. Instead, it is primarily concerned with how evidence of students' learning from assessments, like examinations, can be collected, coherently, validly and reliably, when making inferences about the cognitive capabilities of students. Exploring different ways of estimating the cognitive demand of individuals is important given that the twenty-first century, or the post-industrial era, is an age when cognitive development will be emphasized (Oakland, 1995).

4.1 Policy with respect to the cognitive demand of Biology Senior Certificate examinations

The CBS (DNE, 1984a, 1984b) described in Chapter 2 gave no specific direction as to how different levels of cognitive demand should infuse the teaching and learning of SC Biology. Instead, the CBS provided a general list of learning objectives and an approach to the syllabus at the beginning of the document, which were expected to apply **across all** the elaborated content students were expected to learn, and on which they would be assessed at the end of Grade 12 (Chapter 2, Figure 2.4). Both the list of learning objectives and the approach to the

[105] Seddon (1978) investigated the properties of BTEO and concluded that different taxonomies served different purposes and that no taxonomy was perfect. He argued for comparisons to be made of the extent of the agreement between judges making classifications of test questions using BTEO with that obtained using other taxonomies.

syllabus involve implicit reference to cognitive demand. To avoid repetition, the CBS list of learning objectives and the approach to the syllabus will be listed later in the chapter (Section 4.4) when their relationships to the PET are discussed.

The CBS was the document which determined SC Biology examination policy during one period of this study, namely, 1994 to 2000. In 2001, the revised the CBS examination requirements for government schools (see Chapter 2.1.4). The IEB, a non-government examining body, continued to follow the CBS. Both the CBS and the modified CBS required that SC examinations have specified proportions of, higher order cognitive skills (HOCS)[106] and lower order cognitive skills (LOCS) (Figure 4.2 [A, B]; Figure 4.3 [A, B]).[107] The CBS did not define what LOCS and HOCS were, but the modified CBS and the IEB had different interpretations of what constituted HOCS in terms of BTEO (Figure 4.3 [A, B, C]). Neither the modified CBS, nor the IEB provided explanations for their particular interpretations of the meaning of LOCS and HOCS.

The CBS required that each examination should have approximately 40% HOCS questions for HG and 20% - 25% HOCS questions for SG (Figure 4.2 A). Although the CBS did not define what constituted HOCS it could be inferred from the CBS text that "higher skills" referred to skills other than the mere recall of knowledge, since 60% of the marks in HG and 75% - 80% of the marks in SG are awarded for "recall of knowledge" (DNE, 1984a, p. 2, 1984b, p. 2), and the remainder of the marks (40% for HG and 20 - 25% for SG) were apportioned to "higher skills" (Figure 4.2, A). The author was unable to track how each of the different examining bodies interpreted the CBS requirements with respect to cognitive demand because during the period of post-1994 transformation, when examining bodies were being dissolved and new ones were created, copies of many documents were "lost".[108] Of the syllabus documents that could be traced, the CED (Provincial Administration of the Cape of Good Hope, 1987), the WCED (WCED, 1995a, 1995b, 1996) and HOR (House of Representatives Department of Education, 1986) all retained the wording of the original CBS with respect to HOCS. That is, no reference was made to any particular taxonomy by which to define or recognize different levels of cognitive demand. The HOD examining body used a modification of the BTEO levels of cognitive demand to explicate what was meant by "higher skills" (Isaac, 1990). Bloom's Taxonomy recognizes six categories of cognitive demand: Knowledge (simplest level),

[106] The CBS used the term "higher skills" and the modified CBS used the term "higher order intellectual abilities", "higher abilities" and "insight" – to embrace these different terms the collective phrases HOCS is used, and LOCS is used to convey relevant contrast.

[107] The CBS and the modified CBS were considered to be the minimum requirements for examining bodies, to set examinations (See Chapter 2.1.5).

[108] It is unclear if copies of syllabus-related documents have been mislaid, misfiled or lost from national DoE records.

Comprehension, Application, Analysis, Synthesis and Evaluation (most complex level), and will be discussed in detail in Section 4.2.2. The three categories of HOCS, and their relative weightings, recognized by the HOD were Comprehension (25% HG, 20% SG), Application (10% HG, 5% SG), and a combined category of Analysis, Synthesis and Evaluation (5% HG, 0% SG) (Isaac, 1990). Thus, the HOD, *using their own interpretation* of what constituted HOCS, retained the spirit of the CBS – that 40 % of HG and 20 - 25% of SG examination questions should be above the level of mere recall for each question paper.

A. **Core Biology Syllabus** (DNE, 1984a, p. 2; DNE, 1984b. p.2),

Higher Grade

4.5 Approximately 60% of the marks allocated to the questions in the written examination will be awarded for recall of knowledge. The remaining 40% will be awarded for higher skills.

Standard Grade

4.5 Approximately 75% of the marks allocated to the questions in the written examination will be awarded for recall of knowledge. The remaining 25% will be awarded for higher skills.

B. **Modified Core Biology Syllabus** (DoE, 2001a, pp. 3-4; DoE, 2002b, pp. 2-3)

Higher Grade

3. Each section of BOTH papers will test recall, comprehension and higher intellectual abilities with the following approximate weighting:

Recall and Comprehension : 60%
Higher Order Intellectual Abilities : 40%

Standard Grade

3. Each section of BOTH papers will test recall, comprehension and higher intellectual abilities with the following approximate weighting:

Recall and Comprehension : 75% to 80%
Higher Order Intellectual Abilities : 20% to 25%

Figure 4.2 Weighting of cognitive demand in examinations as stipulated in two SC Biology policy documents.

A. Core Biology Syllabus (DNE, 1984a, p. 2; DNE, 1984b. p. 2)

No prescribed taxonomy for the classification of higher order and lower order cognitive skills

B. National Guidelines for Biology (DoE, 2001a, p. 6; DoE, 2002b, p. 5)

ABILITY LEVELS OF ASSESSMENT FOR THE GRADE 12 EXAMINATION

ADAPTATION OF BLOOM'S TAXONOMY OF OBJECTIVES

Bloom's Categories		Simplified categories	
1. Knowledge	→	Knowledge	A
2. Comprehension	→	Comprehension	B (I; V; N)
3. Application	→	Application	C
4. Analysis 5. Synthesis 6. Evaluation	}	Higher abilities	D

TABLE OF IDENTIFICATION CRITERIA

CATEGORY NAME	CATEGORY REFERENCE	ITEM RECOGNITION DETAILS	WEIGHTING % HG	WEIGHTING % SG
KNOWLEDGE	A	Items merely assessing the recall of facts		
COMPREHENSION	B	Items requiring more than ñÀò and assessing understanding of routine and familiar material	60	80
• **Interpretive**	(BI)	e.g. [,] from verbal to symbolic and/or from symbolic to verbal		
• **Verbal**	(BV)	e.g. [,] explanations		
• **Numerical**	(BN)	e.g. [,] standard exercises		
APPLICATION	C	Items requiring the application of abstractions to new, novel or unfamiliar situations	40	20
HIGHER ABILITIES (INSIGHT)	D	Items requiring: • Analysis of data and pattern recognition • Synthesis of data • Evaluation of data against given criteria		

C. Independent Examinations Board (Requirements for Higher Grade 2001 to 2007, James Buchanan, personal communication, 7 April 2010)

Weighting of cognitive levels	60%			40%		
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation

Figure 4.3 Explanation of cognitive demand categories and requirements stipulated in different SC Biology policies, 2001 to 2007.

In 2001, the modified CBS for national DoE-administered (government) schools stipulated that the HOCS required to be examined for Biology at SC level remain as required by the CBS, that is, 40% for HG and that a range of 20% - 25% be accepted for SG (Figure 4.2 B). In another part of the same modified CBS (DoE, 2001a, 2002b), *exactly* 20% HOCS was stipulated as the requirement for SG (Figure 4.3 [B]). The modified CBS stipulated that HOCS be linked to an adapted version of BTEO (Figure 4.3 [A, B[). In the same year, the IEB also defined their HOCS using BTEO (Figure 4.3 [C[). Bloom et al. (1956) did not categorize each of their six levels of cognitive demand as either LOCS or HOCS. This contrast means that any particular classification of the BTEO levels of cognitive as either LOCS or HOCS is subject to policy makers' interpretations. The DoE, which produced the modified CBS, regarded the BTEO Application level as contributing to HOCS whereas the IEB considered the same BTEO Application level to contribute to LOCS. Neither policy was founded on empirical evidence nor was accompanied by a rationale for their particular conceptualization of LOCS and HOCS. Therefore, who decides what the "correct" HOCS policy might be in the absence of convincing theoretical arguments supporting either? This difference in the LOCS and HOCS requirements of the SC Biology between government policy and that of the IEB has persisted beyond 2007 (Figure 4.4).

A. Department of Education (DoE, 2007g, p. 10)

Weighting of cognitive levels	30%	20%	30%	20%		
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation

B. Department of Education (DoE, 2009b, p. 4)

Weighting of cognitive levels	20%	40%	30%	10%	
	Basic knowledge	Comprehension	Application	Synthesis	Evaluation

C. Independent Examinations Board (IEB, 2007a, p. 24/1)

Weighting of cognitive levels	60%			40%		
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation

Figure 4.4 Cognitive demand categories and requirements stipulated in different NSC Life Sciences policies, 2008 to 2010.

The implications of these differing policies on the assumed equivalence of SC/NSC examinations, is profound, as is illustrated in the example which follows. Based on their Subject Analysis Report (Umalusi, 2010) Umalusi (2011, no page number) justified a downward adjustment of the Department of Basic Education (DBE) (formerly national DoE) 2010 NSC Life Sciences marks because

> [t]he qualitative evaluation of the papers indicated a leaning towards easier and moderate questions which resulted that the 2010 papers were of a lower cognitive demand than in 2008 and 2009. The papers were less demanding and didn't contain sufficient [A]pplication questions.

My own analyses of the DBE NSC Life Sciences papers using the version of the BTEO given in the DBE policy (Figure 4.4 [B[) indicated that there were no Analysis, Synthesis and Evaluation questions, but there were Application questions. Using the IEB interpretation of HOCS policy (Figure 4.4 [C]) would have meant that the DBE question papers included no HOCS questions compared to those of the IEB which contained 40% HOCS questions (IEB, 2009). On what basis could the DBE and the IEB 2010 NSC Life Sciences examinations be considered equivalent? Both carried the same certification.

4.2 Measures of cognitive demand

Evolving conceptions about learning continue to challenge how the validation of assessments are viewed and conducted (Moss et al., 2006). Assessments, like examinations, require that inferences be made from student performances about student competence, and the relationship between (latent) competence and (manifest) performance "involves indeterminacy and imperfection" (Andrich, 2002, p. 38). How then might the construct 'competence' or 'mastery' be understood, and explicated, from student performances in examinations?

In Chapter 3, it was argued that content standards and performance standards could provide a lens through which student performance in SC examinations could be viewed and understood. Content standards were conceptualized as having two components, namely, topics and cognitive demand. The performance standards of an examination which give meaning to what constitutes competence or mastery in that examination in turn derive their meaning from the content standards. Is there a common understanding of how best researchers analyzing assessment tasks might recognize, classify or describe the cognitive demand of a task? Schraw and Robinson (2011) claim there is not.

A survey of the literature revealed a number of different tools, not all called taxonomies, used to classify or group assessment (examination) questions according to explicated characteristics which describe different levels of cognitive demand. Measures of cognitive demand varied between disciplines across several characteristics, the purpose for which classifications using the taxonomy were used, how easy the taxonomy was to use, the experience of the user of the taxonomy, and how cognitive demand was defined. Therefore, because each taxonomy tells its own story about student performance, depending on how cognitive demand was viewed, they should not be used interchangeably within and between studies. Finally, if judgments about cognitive demand are made on the basis of a particular taxonomy, or method, there needs to be both intra- and inter-rater reliability between the classificatory judgments made by raters, to ensure both consistency between raters, and reliability within any particular rater. That is, a principled basis on which inferences about the cognitive demand of examinations can be made, is required.

4.2.1 The quest for a way to measure cognitive demand

Bloom et al. (1956) recognized several characteristics that taxonomies of cognitive demand should have in order to be of practical value. These characteristics were: what was to be classified using the taxonomy (behavioral objectives in the case of Bloom et al. [1956]) should be clearly stated; appropriate symbols were needed to precisely describe the categories used to group what was to be classified; a common interpretation of what the symbols should mean and consensus amongst the people who are using the taxonomy about their classifications. This author's experience over many years was that in South Africa it was difficult to achieve consensus about the classifications of the cognitive demand of assessment questions made by Biology examiners, Biology teachers and trainee Biology teachers using BTEO. Difficulties arose because the classification of some questions straddled more than one category recognized by the BTEO, or because some of the descriptions of categories given in BTEO were interpreted differently by raters, and there was no consistent way of identifying 'new' contexts. One explanation that has been offered for the lack of reliability of classifications using the BTEO was the existence of "fuzzy" category boundaries (Moseley, Baumfield, Elliott, Gregson, Higgins, Miller & Newton 2005. p. 52).

The author of this thesis had also been involved in an Umalusi research study (Umalusi, 2007) which used the Revised BTEO (Anderson et al., 2001) to compare examinations from different African countries. Here she experienced that it was difficult to reach inter-rater agreement in terms of the categorization of questions. While doing this research it became clear that, for this study it was necessary to explore different ways of classifying the cognitive demand of assessments.

Using the Bloom et al. (1956) criteria described in the first paragraph of this section for selecting a taxonomy, 'what' was to be classified in this study needed to be clearly defined. In the absence of cognitive laboratories to determine exactly how students answered examination questions or if students answered examination questions in the same way, the 'what' could not be explicitly called cognitive skills.[109] However, inferences about students' cognitive processes can be made from item response patterns (Gierl, 1997). In this thesis, because there was no cognitive laboratory data, it was necessary to find a proxy which could describe and measure the demonstration (or absence) of mastery for a set of implicit or covert cognitive skills from student answer scripts. Therefore the 'what' in this study could only be the types of activities that examination questions expected students to perform, that is, the *performance expectations.* The author compiled a list of possible performance expectations[110] that might appear in science assessments from a number of sources (AAAS, 1993; CCSSO & Wisconsin Center for Educational Research [WCER], 2004; DNE, 1984a, 1984b; DoE, 2001a, 2001b; IEB, 1996; NRC, 1996). The performance expectations are presented later in this chapter when they are used to illustrate the categories of cognitive demand which they will represent in this study (Section 4.3.2).

Some 13 different classificatory systems, including different taxonomies, were found in the literature. In chronological order the classificatory systems summarized are BTEO (Bloom et al.,1956); classifications of performance expectations used in the 1995 Trends in International Mathematics and Science Study (TIMSS) (Robitaille et al., 1993); an adaptation of BTEO (Zohar, et al., 1998); levels of student understanding recognized in the SOLO taxonomy (Biggs,1999); Webb's Depth of Knowledge categories (Webb, 1999); classifications of performance expectations in the 1999 TIMSS-Repeat (TIMSS-R) (Martin, Mullis, Gonzalez, Gregory, Smith, Chrostowski, Garden & O'Connor, 2000); the Revised BTEO (Anderson et al., 2001); the types of knowledge and use of types knowledge for science assessments (Li, 2001); A New Taxonomy of Educational Objectives (Marzano, 2001); cognitive demand categories for science used in the Surveys of Enacted Curriculum (SEC) model (CCSSO & WCER, 2004); classification of the cognitive domain used in the 2003 TIMSS (Reddy, 2006b); The New Taxonomy of Educational Objectives (Marzano & Kendall, 2007), and an extrapolation of BTEO (Crowe, Dirks & Wenderoth, 2008). Summaries of twelve of these different frameworks (BTEO is given separately in Section 4.2.2) can be found in Appendix 4.1. The author

[109] Messick (1989a) stressed the need to examine extent to which assessment tasks were eliciting the intended cognitive processes in validity studies. Kane (2001), Gierl (1997) and Leighton, Gierl, and Hunka (2004) demonstrated a mismatch between examiners' ideas about what they were testing with respect to cognitive processes, and how students were responding to test items.

[110] This list is considered a work in progress, additional performance expectations may need to be added if the PET is used beyond this thesis.

attempted to classify the performance expectations identified for this study using all these classificatory and was unable to achieve discrete, unambiguous groups of performance expectations, which were considered necessary for the comparative aspects of this study. This outcome was not surprising, given the difficulty that examiners have in the prediction, and the justification for their prediction, of the level of "thinking"[111] required by a question (Black, 1998, p. 66). The search for a suitable instrument to measure cognitive demand for this study was thus continued.

The choice of measure used to determine the cognitive demand of assessments in a study is dependent on how the outcome of the study is to be used. For example, Conley (2003), Fonthal (2004) and Brown and Conley (2007) uses data generated by Marzano's scale of cognitive demand (Marzano, 2001). They used Marzano's classification method because they believed is was best able to represent the higher mental processes important for post-secondary study, and their studies were concerned with determining the alignment between Mathematics and English high school assessments and higher education expectations in some USA states. In her study of the alignment between a battery of science tests developed for Iowa, USA, and Iowa's science standards, Larson (2003) used a modification of Webb's (1999) and the BTEO (Bloom et al., 1956) to determine cognitive demand, because she was not primarily concerned with university expectations.

Given that in more recent years the policy documents prescribing cognitive demand levels of Biology assessments at school level in South Africa had made use of BTEO and that this instrument was the taxonomy with which that the community of Biology educators was most familiar, it seemed logical that, if possible, the BTEO should be modified so that it could more reliably used in this study. Pursuing a modification of BTEO was also consistent with the intention of the original authors of BTEO who envisaged that it would evolve as it was being used, and adapted depending on the discipline and context (Bloom et al., 1956; Anderson et al., 2001). The exercise of trying to use each the 13 classifications named above, indicated how differently suited each framework was to the classification of the performance expectations derived for this study, and if the BTEO was to be modified, how the "new"[112] taxonomy might look. Insights derived about necessary features of a new taxonomy were that a) a separate category was needed for the routine procedures associated with science; b) categories of Application were very dependent on the context of the question and how the context related to a student's experience; and c) that different kinds of knowledge exist.

111 Later, "thinking" is used synonymously with understanding (Section 4.3.1.3).

112 "There is no such thing as a "new" taxonomy [of cognitive demand]; all the likely taxonomies have been invented, and in nearly infinite variety" (Shulman, 2002, p.38). The term "new" is used in this thesis to describe the PET which evolved through modification of BTEO.

The possibility of a two-dimensional framework for cognitive demand to accommodate different kinds of knowledge (Anderson et al., 2001; Marzano, 2001; Marzano & Kendall, 2007) was not pursued because a two-dimensional relationship between topic and cognitive demand had already been envisaged for content standards (Chapter 3). Content standards, as they are used in this study, recognize that different topics can be learned at different levels of cognitive processing. One of the taxonomies, the SEC model (CCSS0 & WCER, 2004) could accommodate all the performance expectations identified for this study because it had a separate category for routine procedures. The next step was to find a way to marry the BTEO and the SEC model of cognitive demand, within current theoretical models of cognitive demand. It is out of this process that the PET emerged. Before the development of the new taxonomy is described, aspects of the history of BTEO, how it has been used to classify behavioral objectives, and some of the criticisms that have been leveled at it over time, are discussed.

4.2.2 Bloom's Taxonomy of Educational Objectives

While the BTEO was developed as a result of the desire of a group of North American college and university examiners to share examination questions, the authors foresaw its value as "[c]urriculum builders" (Bloom et al., 1956, p. 2). The authors were guided by a "vision of what constitutes education for productive learning" (Rohwer & Sloan, 1994, p. 62) and the importance of learning at higher cognitive levels (Anderson et al., 2001; Bloom, 1974; Shulman, 2007). The BTEO provided a standard vocabulary which could be used to describe what any educational objective, and therefore any question, was intended to measure within the cognitive domain. While BTEO was never intended to offer more than a framework to be developed for use within different disciplines or fields according to their specific needs, it has became one of the best-known ways of understanding student learning (Anderson et al., 2001; Marzano, 2001; Marzano & Kendall, 2007) and academic performance (Shulman, 2007). The BTEO has been used in studies at all levels of education, from pre-school education (e.g., Bogan & Porter, 2005) and high school education (e.g., Zheng et al., 2008), to higher education (e.g., Allen & Tanner, 2002; Granello, 2001). Half a century after its development, the BTEO is still used in research that spans a range of different disciplines (e.g., Andrich, 2002; Domin, 1999; Lord & Baviskar, 2007; McConnell, Steer, & Owens (2003); Whiteley, 2006; Zheng et al.,2008), and different countries (e.g., Andrich, 2002; Karamustafaoğlu, Sevįm, Karamustafaoğlu, & Çepnį, 2003; Malan, 2000; Riazi, 2010; Wang & Farmer, 2008) As the BTEO is not linked to any specific discipline or specific educational context, it needs to be clearly (re)contextualized each time it is used (Andrich, 2002).

The BTEO categorized test items within the cognitive domain as either *Knowledge* (least complex), *Comprehension*, *Application*, *Analysis*, *Synthesis* or *Evaluation* (most complex). Specific verbs of behavior were given to explain the evidence necessary to recognize and categorize an educational objective or test item (Bloom et. al., 1956) according to one of the six categories. Collectively, the categories of BTEO were described as "a relatively concise model for the analysis of educational outcomes in the cognitive area of remembering, thinking and problem[-]solving" (Bloom et al., 1956, p. 2). The six categories of cognitive process defined by the BTEO were empirically confirmed by Smith (1970) and in a number of studies noted in Willson (1973). Klein (1972) found that all of the six categories, other than Application, were elicited among a study of tests for seven to nine-year-olds. Using BTEO, Stedman (1973) observed more significant differences when comparing the test scores of high school Biology students for questions at the Comprehension level with questions from the Application level, than he observed when comparing Knowledge and Comprehension questions or Application and Analysis questions. Based on these results, Stedman (1973) questioned the relationship between Comprehension and Application in the BTEO and suggested that further research was needed to specifically investigate this relationship.

The BTEO made another assumption, namely, that a cumulative hierarchy existed between the categories Knowledge to Evaluation (Bloom et al., 1956; Bloom et al., 1971). That is, each successive category assumed that a student performing at that level would have all the knowledge and skills required to perform at the preceding level(s) together with what defined that category. Roberts (1976) supported this assumption but suggested that the link between Knowledge and Comprehension was weak and that the category Application was difficult to place. Because the category Knowledge is the least complex level of BTEO, it has been described as "the link between subject matter or content and the Taxonomy" (Kropp, Stoker & Bashaw, 1966, p. 69). Booker (2007, p. 352) contested the foundational position of the category Knowledge in BTEO and wrote that BTEO "fuels the belief that HOCS can exist in isolation from specific content". However, he did not provide empirical support for his position and he failed to acknowledge the cumulative nature of BTEO which makes Knowledge an integral part of all the higher levels of the taxonomy. Bloom et al. (1956, p. 33) themselves stated that "[t]he intellectual abilities represented in the taxonomy assume knowledge as a prerequisite". Furst (1981) described both philosophical and educational objections to the under-representation of specific content[113] in the BTEO. The Revised BTEO (Anderson et al., 2001) acknowledged that knowledge permeates all the other levels of cognitive processes (and vice versa) when they

[113] Content was used by Furst (1981) in a similar way to knowledge in the BTEO.

created a two-dimensional taxonomy with a knowledge dimension and a cognitive process dimension (Appendix 4.1 [I]).

Recent objections have been made against the popular depiction of BTEO as a pyramid with Knowledge as its base and Evaluation at its apex because this "sends the wrong message about the importance of knowledge in learning" (Wineburg & Schneider, 2009/2010, p. 57). Representing the BTEO as a pyramid with Knowledge at its base and Evaluation at its apex was not promoted by Bloom et al. (1956). It is an interpretation of the BTEO that was made popular by subsequent users of the taxonomy. It is difficult to conceive of a relationship between the categories of the BTEO resembling the upside down pyramid of Wineburg and Schneider, which still retains Knowledge as the base of the pyramid. However, if one accepts that Knowledge is the base of the BTEO and that the BTEO is cumulative (Bloom et al., 1956), and if a pyramid shape is envisaged for the BTEO, the widest part of the pyramid would be the category Evaluation since it embraces all the categories below it.

Different categories of studies of the BTEO appear in the literature—the taxonomy's own assumptions and categories have been examined (e.g., Madaus, Woods, & Nuttall, 1973; Stoker & Kropp, 1964), questions have been asked about how consistently raters using the BTEO concur on their classifications (e.g., Moseley et al., 2005, Stoker & Kropp, 1964; Wood, 1977), and broader analyses of the taxonomy with respect to learning theories have been undertaken (e.g., Rohwer & Sloan, 1994). A detailed discussion of the literature in connection with the BTEO is beyond the scope of this thesis. What follows instead is a synthesis of some of the relevant research concerning BTEO which influenced the construction of the 'new' taxonomy which emerged out of BTEO for this study.

Kropp, Stoker and Bashaw (1966), Smith (1968, 1970)[114] and Stoker and Kropp (1964) claimed to have empirically supported the hierarchical structure of the taxonomy but suggested that, for science, the order of Synthesis and Evaluation be reversed, or that they be placed parallel with each other. Hill and McGaw (1981) found that a re-examination of the Stoker and Kropp (1964) data separated the category Knowledge from the remaining five categories which were arranged in a hierarchical way. Also citing empirical evidence, Madaus et al. (1973) observed a break in the assumed taxonomic hierarchy and proposed a Y-shaped structure where the stem of the Y went from Comprehension to Application. One branch of the Y went from Application to Analysis and the other branch went from Application to Synthesis and Evaluation, which were also the two categories least dependent on integration with the lower levels. Kunen, Cohen

[114] Seddon (1978) contested the results of these studies on statistical grounds.

and Solman (1981) empirically tested the cumulative hierarchical assumption of BTEO and reported only moderate support for the hierarchy. Kunen et al. (1981) also found that the category Evaluation was misplaced at the apex of the hierarchy. In the light of these and other challenges to the assumption of the hierarchical nature of BTEO (Kreitzer & Madaus, 1994) it was suggested that it be considered to "resemble a hierarchy" Rohwer and Sloane (1994, p. 47) where relationships between the categories BTEO and therefore the structure of the hierarchy may be domain-specific (Nordvall & Braxton, 1996).

While Bloom et al. (1956) and Krathwohl et al. (1964) sometimes regarded complexity and item difficulty as synonymous in BTEO, these terms are distinct attributes that are often not even correlated (Hancock, 1994).[115] Marzano (2001, p.10) challenged that these terms could be used synonymously, citing "the well-established principle in psychology that even the most complex of processes can be learned at the level at which it is performed with little or no conscious effort". "The more familiar one is with a process, the more quickly one executes it and the easier it becomes" (Kendall, Ryan and Richardson, 2005, p. 5). Kropp et al. (1966, p. 74) suggested that what was lacking was a way of operationally distinguishing the two terms and proposed that

> complexity might be determined by logical analysis of the operations which apparently should be performed and [have] to be performed [to get to an answer] and difficulty might be defined as the proportion of a group who respond to the item correctly in the absence of operational definitions which distinguished them from one another.

Sousa's (2006) distinction between complexity and difficulty is more straightforward and regarded complexity and difficulty as being separate but related. He explained complexity as referring to the different levels of cognitive demand, defined by BTEO, and difficulty as the amount of effort that a person needs to expend within each level of complexity to perform a learning task. Separating difficulty and complexity has been further complicated because the difficulty of questions is rarely controlled for in studies (Crooks, 1988), which makes it awkward to determine the relationship between student performance with respect to difficulty or complexity of questions.

The authors of the BTEO described their category Knowledge as requiring the lowest cognitive demand, and Evaluation the highest cognitive demand, but did not explicitly assign the other categories to one or the other of these groupings. Consequently, many studies using BTEO were inconsistent in how the six BTEO groups comprised each of the higher order and lower order questions were defined. For example, the DoE and the IEB examining bodies

[115] Other authors considered complexity as only one component of cognitive demand (Dall'Alba & Edwards, 1983; Edwards & Dall'Alba, 1981; Sousa, 2006).

each had different interpretations about categories of BTEO which comprised higher order cognitive demand within South African SC Biology examination requirements (Section 4.1). Other examples, from the literature of different interpretations of BTEO with respect to cognitive demand will be discussed in Section 4.2.3.

Another criticism of BTEO is the role of understanding in the taxonomy. Ormell (1974, 1979) criticized the BTEO because it included 'understanding' as a constituent part of the taxonomy rather than as a collective descriptive term, which could have been better used to incorporate some or all of the other categories. The position of understanding in a taxonomy of cognitive demand will be further discussed in Section 4.3.1.

Bloom et al. (1956) claimed to have achieved a high level of inter-rater agreement between users of the BTEO. This view was supported by the research of Stoker and Kropp (1964) and Gierl (1997) but was contested by the work of Baughman and Mayrhofer (1965) and Seddon (1978) which showed that inter-rater reliability varied considerably.[116] Kreitzer and Madaus (1994) recommended the training of raters to obtain better consensus between the classifications of raters using BTEO. Suggestions have been made that inter-rater agreement fails when a questions falls into one of the higher BTEO categories, that is, Analysis, Synthesis or Evaluation (Nordvall and Braxton, 1996).

Black (1998, p. 65) considered "all that survives such empirical investigation [about BTEO] is that there is a broad distinction between Knowledge on the one hand and higher order skills represented by Synthesis and Evaluation on the other". This section described some of the empirical challenges that have been made to BTEO. Section 4.2.3 below documents a series of inter-referenced studies demonstrating how the BTEO has been misrepresented in studies.

4.2.3 An example of misinterpretation and misrepresentation of Bloom's Taxonomy of Educational Objectives

This section demonstrates, by tracking the references cited in arguments made in two recent and connected publications, Freeman, Haak and Wenderoth (2011), and Haak, HilleRisLambers, Pitre and Freeman (2011) (Figure 4.5). These examples will highlight two problematic ways in which researchers have inconsistently measured and discussed cognitive demand. First, researchers have, without empirical evidence, quantified levels of the BTEO;

116 At the beginning of their study, Baughman and Mayrhofer (1965) recorded an inter-rater agreement of 64% which increased to 84% at the end of their study.

	Categories					
Bloom et al. (1956)	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
	Knowledge	Intellectual abilities and skills				
	Less complex →					More complex
Krathwohl (1964)	Remembering information	Problem-solving				
Zoller (1993)	Low[er]-order cognitive skills ï knowing / understanding		Higher-order cognitive skills - decision making, problem-solving, creative/critical thinking			
Lord (1997)	Memorized / rote information	Understanding of material ï its comprehension and implication				
Allen & Tanner (2002)	Foundational		More cognitively complex			
	Less complex →					More complex
Krathwohl (2002) [a]	Remember	Understand	Apply	Analyze	Evaluate	Create
Bissell & Lemons (2006)	No critical thinking		Critical thinking ï higher order skills			
Zheng et al.(2008)	Lower-order thinking		Higher-order thinking			
Crowe (2008)	Lower-order cognitive skills			Higher-order cognitive skills		
Freeman et al. (2011)	Lower-order cognitive skills		Higher-order cognitive/thinking skills			
Haak et al. (2011)	Lower-order cognitive skills		Higher-order cognitive skills			

Note:
a Krathwohl (2002) described a second dimension, the knowledge dimension, which is not shown here, but is a part of the Revised BTEO (Anderson et al., 2001).

Figure 4.5 Categories of cognitive demand recognized by various authors which are directly and indirectly referenced by Haak et al. (2011) and Freeman et al. (2011). Terminology is that used by each author(s).

Second, researchers have moved between instruments which are conceptually different, especially with respect to what constituted LOCS and HOCS. These problems are clarified below, and each is tackled by starting with the most recent publications.

Haak et al. (2011) concluded, in their study drawing on the results of Freeman et al. (2011), that regular practice of HOCS could reduce the achievement gap between advantaged and disadvantaged students in introductory university biology classes. Freeman et al. (2011) and Zheng et al. (2008) which they cite, ordered the BTEO categories from 1 (Knowledge) to 6 (Evaluation), and made further calculations which weighted the BTEO categories accordingly. Based on these quantified BTEO weightings, Zheng et al. (2008) pronounced about the comparability of college placement and medical school admission examinations with respect to HOCS and Freeman et al. (2011) concluded that the increased emphasis on HOCS in the course design of introductory biology courses resulted in improved student performance. Both studies are seriously flawed because they make an empirically unsupported assumption that, for example, Evaluation is 'worth' six times more than Knowledge. Despite this both Zheng et al. (2008) and Freeman et al. (2011) were published in one of the most prestigious science periodicals, *Science.*

The problems associated with researchers moving interchangeably between conceptually different interpretations of LOCS and HOCS are more complex. Haak et al. (2011) referenced Crowe et al. (2008) as a standardized framework for assessing levels of learning but they did not use this framework. Haak et al. (2011) used the results of Lord (1997), which make no reference to BTEO to corroborate their findings and those of a previous article (Freeman et al., 2011) which used a particular interpretation of what constitutes HOCS using BTEO (Figure 4.5). Lord (1997) included comprehension/understanding with HOCS, and concluded that active learning influenced understanding but not memorization. Despite placing Comprehension in LOCS, Haak et al. (2011) and Freeman et al. (2011) used Lord's (1997) argument, inappropriately, to reinforce their conclusion that active learning had no influence on Comprehension, but had an influence on levels Application and above.

Freeman et al. (2011, p. 179) quoted Bloom et al. (1956) and Krathwohl (2002) as having identified "six levels of understanding and any topic". This view is not correct. The BTEO first level is not understanding – the remaining five categories can be, if one accepts that BTEO is a cumulative hierarchy subsuming Comprehension (understanding) into every level above it. Similarly, in the Revised BTEO (Krathwohl, 2002) is the Understand is the second of six categories of cognitive demand. Freeman et al. (2011) and Haak et al. (2011) made reference to the use of the Crowe (2008) interpretation of LOCS and HOCS (Application is both LOCS and of HOCS), while simultaneously defining Application as part of HOCS. Haak et al. (2011) noted that most of the questions were of the category Application yet it is not clear how they treated Application in their further analyses (i.e., as LOCS or HOCS) and therefore how the results reported might be affected.

Crowe et al. (2008) described their study as evolving out of that of Zheng et al. (2008) and referenced Zoller's (1993) categories knowing and understanding as being synonymous with Knowledge and Comprehension of BTEO. Zheng et al (2008) quoted Bissell and Lemons (2006) and Allen and Tanner (2002) and the original BTEO and the Revised BTEO. In their study Zheng et al. (2008, p. 414) treated the original BTEO and Revised BTEO as the same: "we used Bloom's Taxonomy (7-9) to quantify the level of learning that students are asked to demonstrate on a sample of course exams and standardized tests" . References 7 to 9 were: Bloom et al.. 1956; Krathwohl, 2002; Anderson and Sosniak (1994). In discussion of their results, Zheng et al (2007) used the original BTEO, and Bissell and Lemons (2006) and Allen and Tanner (2002) who conceptualized LOCS and HOCS differently (Figure 4.5).

Bissell and Lemons (2006) separated the BTEO categories only the basis of whether critical thinking is involved or not. Krathwohl (2002) renamed the BTEO categories but does not assign LOCS and HOCS to any category. Allen and Tanner (2002, p. 64) acknowledged the BTEO levels of complexity and allowed educators to make "a distinction between lower-level [Knowledge and Comprehension] and higher-order knowing and thinking (commonly referred to as critical thinking)". Lord (1997) made no reference to BTEO. What he said was that students who are constructivist-taught do better in assessments requiring an understanding of the material than do traditionally-taught students. There was no discernable difference in the performance of traditionally-taught students and constructivist-taught students in assessments requiring a rote information or memorized response (Lord, 1997). Zoller (1993) makes reference to BTEO in text, but neither cites nor references it at the end. In their foreword to the Taxonomy of Educational Objectives – Affective Domain, Krathwohl, Bloom and Masia (1964) described two similar mega-groups as in the BTEO with slightly different names, remembering information and problem-solving (Figure 4.5).

The above examples of how taxonomies of cognitive demand have been confusingly used, emphasizes how researchers have assumed that taxonomies, and how they are interpreted conceptually, can be used interchangeably and that they measure the same 'thing'. This confusion indicates a need for an explicitly and unambiguously worded taxonomy of cognitive demand for this thesis – PET.

4.3 The Performance Expectations Taxonomy

This section describes the development of an instrument, called the PET, which is used to extract information about the performance expectations of the examination questions. Specifically, PET will provide information about the cognitive demand component of the content standards generated from examination question papers and reported in Chapter 6.

Kreitzer and Madaus (1994) recommended the training of raters to get better consensus between the classification of raters using BTEO. Therefore, inter-rater reliability of PET users needed to be established. A number of authors, including Bloom et al. (1956), and Norvall and Braxton (1996), and Andrich (2002) made reference to the importance of prior learning that students bring to a question which they are answering and how this variable would influence the way in which a question is classified with respect to cognitive demand. For example, for a student who had been taught and who had learned the solution to a particular question, recalling the solution would make the question one of knowledge. Another student, who had not been taught or had not learned the solution to the question might have had to arrive at the answer by, for example, application or a process of trial and error. Kropp et al. (1966, p. 71) suggested two ways in which the prior-knowledge could be addressed: by assuming that students have “the content available to them by virtue of common antecedent experiences” or by ensuring that all students have had access to the same content before they are tested, so that at the time of the test new content is “equally unfamiliar to all students”.

The PET takes prior knowledge into account by using the factual, conceptual and procedural knowledge explicitly stipulated in the syllabus documents (Chapter 2.1.4) to define the minimum that students were expected to be taught and to learn, that is, the intended curriculum. Anecdotal evidence from discussions with Biology teachers suggests that not all SC Biology candidates were exposed to an enacted curriculum that mirrored the intended curriculum for Grade 12 Biology. In the absence of empirical evidence to support or refute such suggestions, this study makes an assumption that all students had access to the same intended curriculum as was conveyed by the syllabus document. This study acknowledges, but does not take into account, that more experienced teachers might have exposed their students to skills beyond the minimum recognized in this study. These two assumptions are declared because the PET uses the SC intended curriculum (prior knowledge) to distinguish between the different categories and between LOCS and HOCS.

4.3.1 Description of the Performance Expectations Taxonomy

In the social sciences, taxonomies are used to sort the objects of study (i.e., the performance expectations of a question) into groups based on the similarities and differences between the objects so that each group is as homogenous as possible and as different as possible from other such groups (Moseley et al., 2005). Given that "taxonomies are primarily descriptive devices and do not offer causal explanations" (Moseley et al., 2005, p. 38), this section describes the groups, called categories, so as to follow the terminology of BTEO of performance expectations conceptualized in the PET instrument developed for this study. The relationships between the categories which result from the particular ways that the categories are conceptualized are also discussed. Like BTEO, which incorporated no single particular learning theory (Bloom, et al., 1956; Anderson et al., 2001; Andrich, 2002), the concept of cognitive demand—and the PET subsequently developed in this study—drew on ideas from a number of different researchers about how people understand and learn. Given that a single theory of learning has yet to be found (Anderson et al., 2001), different theories about how people learn, think, understand and solve problems are incorporated into the discussion where appropriate. The PET also sought a marriage between BTEO and the SEC model of cognitive demand (Section 4.2.1).

The PET recognizes five categories of performance expectations: *Memorize*, *Perform-Routine - Procedures*, *Explain, Analyze* in familiar contexts and *Apply* in new contexts (Figure 4.6).[117] A dichotomous key developed to aid the classification of questions,[118] during the validation of the PET was found to be extremely useful in achieving inter-rater agreement (Section 4.3.2) (Figure 4.7). The wording of the key and the table used in the PET validation process (Appendix 4.2 and Appendix 4.3) differs slightly from those given in Figure 4.6 and Figure 4.7, which are used in the analysis of examination question papers (Chapter 5).

The five PET categories can be grouped in two different ways. One way of grouping the categories separates the two categories that require no demonstration of understanding (i.e., Memorize and Perform-Routine-Procedures) from the three categories that require a demonstration of understanding (i.e., Explain, Analyze in familiar contexts and Apply in new contexts)—these mega-categories are called *Rote-And-Routine* and *Demonstrate-Understanding*, respectively (Figure 4.6). Another way in which the five categories can be grouped is according to what the PET recognizes as *LOCS* (i.e., Memorize, Perform-Routine-

---

117 The performance expectations are numbered for convenience rather than presupposition of order or hierarchy.

118 Called scorable events from Chapter 5 onwards.

Procedures, and Explain) and *HOCS* (Analyze in familiar contexts and Apply in new contexts) (Figure 4.6).

What follows are descriptions of the five categories of the PET (Section 4.3.1.1), together with the rationale for two different ways in which the PET categories can be grouped (Rote-and-routine vs Demonstrate-Understanding; LOCS vs. HOCS) (Section 4.3.1.2). The relationships between the five categories are also discussed and demonstrated using examples of questions taken from the SC Biology examinations analyzed for this thesis. Discussion of some of the current theories about how cognitive demand is viewed in the literature form the basis for how both the PET categories and their groupings are conceptualized, and explicated, as well as how they can be recognized.

4.3.1.1 Categ*ories of performance expectations*

The descriptions of the five categories of the PET, that is, *Memorize* (A), *Perform-Routine-Procedures* (B), *Explain* (C)*, Analyze* (D) and A*pply* (E), are given from the perspective of analyzing questions, and the type of answers which they require, given to students in the SC Biology examinations. These descriptions of the five PET categories should be read in conjunction with Figure 4.6 and Figure 4.7.

If the PET were to be used for the analyses of the tasks which comprise formative assessments during the year, the context would need to be redefined to ask "if the students doing the task have or have not been taught the content or not?". Such re-contextualization would be necessary because the PET uses what the student was (required to have been) taught to distinguish between LOCS and HOCS (Section 4.3.1.3).

A. ***Memorize knowledge***

The answers to a memorize question "only" [119] indicate whether a student is able to recall or recognize science facts, terms, definitions, concepts or formulae which the CBS required to be taught or which were given as a part of the question. It is vital that "only" be a part of the description of this category, because a memorize question does not ask a candidate to demonstrate that she/he can do anything *other than* recall or recognize information. Such a question is related to the *retention* of knowledge which has been associated with cognitive

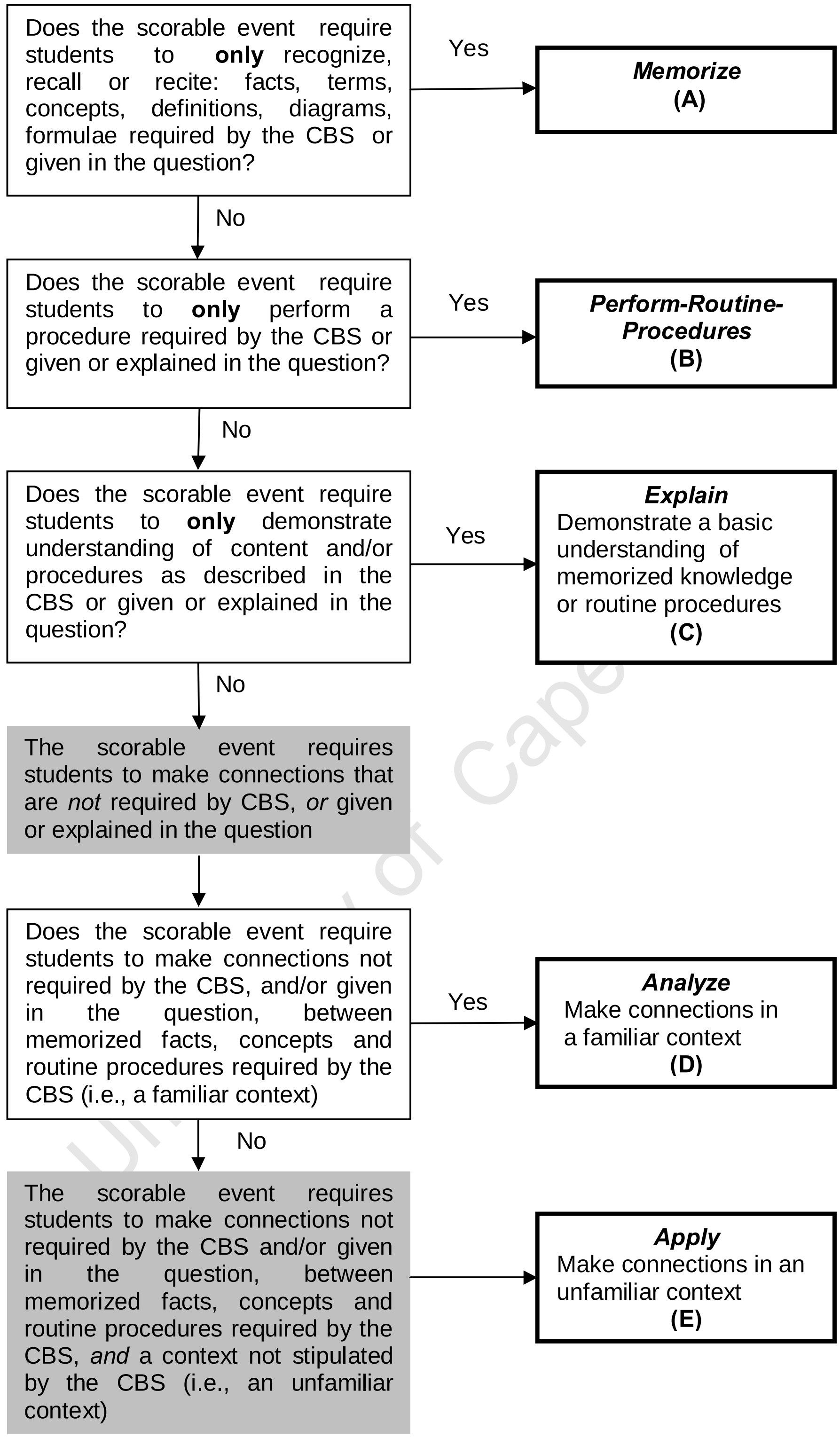
[119] "Only" is used in quotes to indicate that the question requires a student to demonstrate nothing more than what is described by that specific category.

<table>
<tr><td colspan="3">LOWER-ORDER COGNITIVE SKILLS

Content, procedures or connections explicitly listed in the syllabus or given in the question</td><td colspan="2">HIGHER-ORDER COGNITIVE SKILLS

Use learned content, procedures or make connections in ways not explicated in the syllabus or given in the question</td></tr>
<tr><td colspan="2">ROTE-AND- ROUTINE (i.e., no demonstration of understanding required)</td><td colspan="3">DEMONSTRATE-UNDERSTANDING (i.e., using acquired knowledge and skills)</td></tr>
<tr><td colspan="2">ACQUIRE important information and skills</td><td>MAKE MEANING of important information and skills</td><td colspan="2">TRANSFER meaning t o new situations</td></tr>
<tr><td>A
Memorize (i.e., knowing)</td><td>B
Perform-Routine-Procedures (i.e., doing)</td><td>C
Explain – demonstrating a basic understanding of memorized knowledge and routine procedures</td><td>D
Analyze information / make connections, not required by the syllabus or given in the question, using memorized knowledge and routine procedures in familiar contexts (i.e., of the syllabus or given in the question)</td><td>E
Apply (use) concepts / analyze information / make connections in new contexts (i.e., outside of syllabus and not given in the question)</td></tr>
<tr><td>1. Recall / recognize science terms, facts, definitions, concepts
2. Recall / recognize scientific formulae</td><td>3. Make measurements
4. Make a scientific drawing
5. Make observations / describe objects, processes, results
6. Read values / information from graphs
7. Compute
8. Use given formulae
9. Use / assemble / handle appropriate tools, apparatus
10. Conduct routine / explained experiments
11. Test the effects of different variables in routine experiments
12. Collect and record data
13. Organize and display data in Tables / graphs / charts as instructed
14. Neatness and presentation of work</td><td>15. Explain / show understanding of learned concepts / routine procedures / processes</td><td>16. Observe and explain student / teacher / given demonstrations
17. Explain methods of science and inquiry
18. Classify and compare data (similarities and differences)
19. Analyze data, recognize patterns / trends
20. Reason inductively / deductively
21. Draw conclusions
22. Identify faulty arguments or misrepresentations of data
23. Generate questions or make predictions from prescribed knowledge or routine procedures
24. Present analyzed information / results</td><td>25. Generate questions / hypotheses or make predictions from unlearned / experimental data
26. Select, use and integrate science concepts / formulae / routines
27. Test the effects of different variables
28. Recognize experimental design errors / appropriate use of controls
29. Synthesize content and ideas from several sources
30. Plan and design an investigation / experiment to address a given / generated problem or question or hypothesis
31. Organize and display data in tables, graphs or charts of own design
32. Reason inductively / deductively
33. Apply and adapt science information to real-world situations
34. Apply science outside the context of science
35. Build or revise a plan / theory
36. Present applied concepts and connections
37. Construct an argument</td></tr>
</table>

Figure 4.6 Categories of the PET and the kinds of performance expectations identified for each category.



Note:

a Syllabus refers to CBS (DNE, 1984a, 1984b) and guideline documents (DoE, 2001a, 2002a, 2005b) .

Figure 4.7 Key to classifying a scorable event according to the PET. Shaded boxes describe distinguishing features of the group(s) that follow and do not require decisions to be made.

processes of recognizing and recalling (Mayer, 2002). This category is embraced by Perkin's (1998, p. 39) phrase "knowledge on tap".[120] As this PET category concerns the "knowing" of required science facts, terms, definitions, concepts or formulae, the intellectual product of this category is known as memorized knowledge.

B. ***Perform-Routine-Procedures***

The answers to a Perform-Routine-Procedures "only" question indicates if a student is able to perform particular procedures, including some psychomotor procedures (Anderson et al., 2001; Bloom et. al., 1956; Marzano, 2001) associated with the "doing" of science and which the CBS required be taught. To be classified in this category, a question requires a student to "only" mechanistically apply taught procedures but not necessarily understand the procedures—which are familiar to the student because they are required by the CBS. Examples of questions in this category would require students to make observations, draw a graph as instructed and read values from a graph, apply given formulae, perform explained experiments and record the results of experiments. This category would be similar to what Anderson et al. (2001) referred to a subcategory named 'Executing' within their category 'Apply', so categorized because the student is required to perform the procedure with little thought. Perkins (1998, p. 39) referred to "routine procedures on tap" as skills.

C. ***Explain (memorized and routine procedures)***

Articulating what it means for a student to understand something is difficult because teachers' interpretations of student understanding are "deeply rooted in assumptions and values that usually remain tacit. Surfacing such tacit knowledge is intellectually difficult and personally revealing" (Wiske, 1998, p. 68), but educators are able to recognize understanding when they see it (Perkins, 1998a). The PET recognizes, like Perkins (1998a), that there are degrees of understanding (Section 4.3.2) and that "[u]nderstanding is a matter of being able to think and act flexibly with what you know" (Perkins, 1998a, p. 42). Recognizing this category as "basic" understanding, implies that other levels of understanding (i.e., Analyze and Apply) will be recognized in the PET.

The answers to a question classified in this category—Explain—demonstrate "only" a basic understanding of memorized knowledge and routine procedures which are specified in the CBS. Questions classified as Explain would require students to explain, or show evidence of

120 Perkins (1998a, p. 39) uses "on tap" to mean that a "student can reproduce it when asked".

understanding, terms, concepts, processes (PET category Memorize) or routine procedures (PET category Perform-Routine-Procedures), required by the CBS. Using a suggestion made by Ramsden (2003, cited by Schwartz, Sadler, Sonnert, & Tai, 2009) that memorization is an active process during which students may need to deconstruct both structures and connections they are attempting to commit to memory, this PET category–Explain–would require students to show evidence of the deconstructions of knowledge which they have made. The "only" is necessary here because the question does not require the student to function outside of the explicit boundaries of biological knowledge as determined by the CBS. In the CBS, the biological knowledge was presented like a 'laundry list' with very few explicit connections made between the items (Chapter 2) within a topic. The CBS made an assumption that the student would be able to use that knowledge in different ways (Chapter 2, Figure 2.4), and therefore that they can make the necessary connections, in this case within a topic.

By recognizing that the category Explain is dependent on what should have been taught and learned, the PET uses understanding in a slightly different way to other definitions of understanding "as the ability to use knowledge in *novel* situations" (Mansilla & Gardner, 1998, p. 182, emphasis added). However, Mansilla and Gardner (1998, p. 182) clarify an early stage understanding as "[f]irst, it requires a shift in focus from isolated facts about the world to broader, richly organized conceptual frameworks of examples and generalizations that are currently accepted as warranted in the domains taught". The PET view of Explain as basic understanding reflects that a question in this category requires a student to demonstrate the isolated facts or skills[121] of the CBS were understood in a broader context than was explicitly stated in the CBS, and in a way that makes sense biologically, and within one of the topics identified in the CBS.

D. ***Analyze (in familiar contexts)***

The answers to questions in this category of the PET indicate whether students are able to break information into component parts to identify how that information fits together beyond the boundaries of topics as stipulated by the CBS. Students are expected to find links between memorized knowledge and routine procedures and to determine which material is relevant and which is extraneous to the task in question. In this PET category, the material for the Analyze task, that is, *all* the component parts, elements, ideas or processes are either stipulated by the CBS or given to the student in the question. The student is assumed to have a basic understanding thereof (PET category Explain) (Figure 4.6) and so the context is considered to

[121] See Chapter 2.1.4 .

be a *familiar* one. The answers to Analyze questions are new in the sense that they would require knowledge and skills that had been stipulated by the CBS or given in the question, to be used in an answer not given as such in the CBS. Because analyze requires the student to make connections beyond the topic(s) in the CBS in which it appears, Analyze is considered to be a deeper level of understanding.

The author felt that it was important to explicate this category as above because the CBS organized the knowledge and routine procedures that students were to be taught into discrete topics, with few inter-topic connections made explicitly. What resulted from this CBS design was that students were not always taught to make inter-topic connections and hence may have missed much of the biological sense of what they were learning. A report by a ministerial committee explaining the results of investigation into the SC examination, quoted a Biology HG examiner as saying " [t]he way of mixing questions from different chapters is of disadvantage to candidates. Questions should be set according to chapter … this will enable candidates to target specific questions that they can score most on, rather than confusing them." (DoE, 1998, p. 11). This comment highlights the dangers associated with assuming that the users of the CBS could/would apply different levels of cognitive demand across a laundry list of knowledge and skills to make the connections between the knowledge and skills that are necessary for a deeper understanding (Vitale & Romance, 2006).

This conceptualization of the category Analyze is consistent with the view of Mayer (2002) who considered analysis to be an extension of understanding and a prelude to higher order learning such as evaluating or creating.

E. ***Apply (in new contexts)***

Answers to questions in the category Apply share similarities with Analyze questions in that the results of the application process are new. The answer to Apply questions would also include knowledge and skills that are not stipulated by the CBS or given in the question. What distinguishes the category Apply in the PET is that *not all* the component parts, elements, ideas or processes required to perform the application task are stipulated by the CBS or are given to the student in the question. That is, the task is *unfamiliar* to the student.[122] The student would need to think beyond the familiarity of the CBS or the question to locate or create the missing components needed to successfully perform the application task. In order to locate or create the missing components they may need to *make connections between what they know*

[122] Some teachers, in some schools may have taught students beyond the CBS.

(PET category Analyze) and what is missing once what is missing has been identified, or they may just need to *connect what they know* (PET category Explain) and what is missing once what is missing has been identified, to answer the question.

4.3.1.2 *Relationships between the categories of performance expectations*

The way that the categories of the PET have been conceptualized implies specific relationships between the categories, all of which are rooted in the facts, terms, concepts, definitions, diagrams, formulae and routine procedures explicitly required by the CBS, or given in the question knowledge (Figure 4.8). Questions from categories Memorize and Perform-Routine-Procedures only require that students demonstrate that they can recall or recognize the required facts, terms, concepts, definitions, diagrams, formulae or perform the required routine procedures─no demonstration of understanding is required. Explain questions require students to demonstrate that they understand the required knowledge and required routine procedures in a familiar context of the CBS. Analyze questions require an extension of the basic understanding required by Explain questions because students need to make connections not required by the CBS, and/or given in the question, between memorized facts, concepts, routine procedures required by the CBS, but still within the familiar context of the CBS. Apply questions requires students to operate within a new context by making connections between what the CBS requires be taught *and* a new context outside of the CBS. Positioning the PET category Apply in this way is consistent with how Miller, Williams and Haladyna (1978) viewed their cognitive level Applying. Miller, et al. (1978) considered four cognitive levels higher than factual recall, that is, Summarizing, Predicting, Evaluating and Applying, and although they did not rank their levels in terms of complexity, they recognized that Applying could include their other categories. Conceptualizing application tasks according to the PET category Apply, addresses difficulties which were discussed in Section 4.2.3 about how the BTEO category Application has been differently interpreted in different studies.

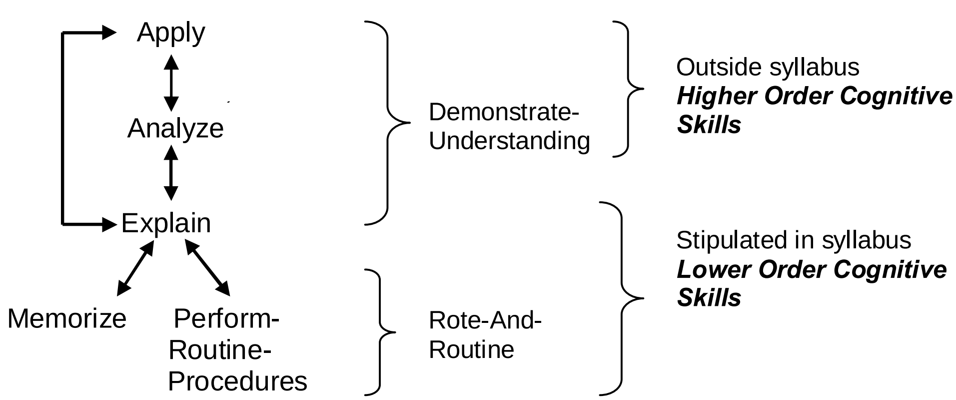


Figure 4.8 Relationships between the categories of the PET.

Wiggins and McTighe (2008) explained learning as when students *acquire* important information and skills (PET Memorize and Perform-Routine-Procedures); *make meaning* of that content (PET Explain) and effectively *transfer learning* to new situations (PET Analyze and Apply) (Figure 4.8). In a study which looked at the role of prior knowledge and reasoning skills in problem solving Chang (2010) concluded that the role of prior knowledge in problem solving was dependent on the whether the problem solving involved convergent- or divergent-thinking. It may therefore be possible that there are alternate routes to the PET category Apply (E).

The Figure 4.8 has deliberately been drawn using double-ended arrows to "challenge the common practice of teaching knowledge and skills for acquisition first and then teaching for meaning and transfer later. Rather, we must recognize that the purposeful and effective use of content is the ever present goal" (Wiggins & McTighe, 2008, p. 41) at all cognitive levels. New knowledge and skills can be learned while engaging in higher level intellectual activities.

4.3.1.3 *Different ways of grouping categories of performance expectations*

Just as there are many different taxonomies of cognitive demand, there are several different ways in which the categories of taxonomies have been grouped together. The author has chosen two different ways to group the five categories of the PET (Figure 4.6) both of which are based on how the CBS policy documents which guide this study articulate the teaching and learning of Biology. The first grouping of the PET categories forms two mega-categories called *Rote-And-Routine* and *Demonstrate-Understanding* (sub-section [a] below) and the second grouping forms two mega-categories called *LOCS* and *HOCS* (sub-section [b] below). The composition of the mega-categories which result from each of these two PET groupings differs only in where the PET category Explain is placed, but the language associated with the names that could have been given to the PET mega-categories is highly varied in the literature.

The mega-categories are named and described according to the different perspectives of various researchers who have written about teaching, learning and knowing in the cognitive demand literature. Increasingly, more authors are emphasizing the need for teaching for understanding and higher order thinking skills, in all disciplines, for all students and at all levels of education systems (e.g., Hirsch, 1996; Perkins, 1998; Raudenbush, Rowan & Cheong, 1993; Silva, 2009; Zohar & Dori, 2003). These views make it important to be explicitly clear about what understanding and higher order skills mean. Discussion of the two ways of grouping the PET categories will illustrate how the same terms, for example, rote or understanding or higher order skills, have subtly different meanings in different contexts, which made it necessary to explicate how such terms are both conceptualized and operationalized in this study.

a. ***Rote-And-Routine vs Demonstrate-Understanding***

A motivation for the recognition of two mega-categories, *Rote-And-Routine* and *Demonstrate-understanding* (signals presence of/ possession of understanding) from the five PET categories came from the CBS which stated that "[t]he approach to the [Biology] course should, as far as possible, embody the following important principles: … constant emphasis should be placed upon facts being understood, interpreted and applied rather than being merely memori[z]ed" (DNE, 1984a, p. i,; 1984b, p. i). In their work on the designing of assessments, Lord and Baviskar (2007) recognized two divisions of questions, namely, those that related to content knowledge and those that related to understanding. Therefore "understanding" used by Lord and Baviskar (2007) clearly means something more encompassing than the "understanding " referred to in the CBS.

Wiske (1998) wrote about the importance of explicit and public articulation of what "it" is that we want students to understand, and Perkins (1998a) argued that understanding varies between topics, disciplines and the experience and development of a student. Therefore, in an explanation of how evidence of understanding and degrees of understanding (Perkins, 1998) would be framed for this study was necessary. It was also necessary to clearly distinguish the group of categories collectively called *Demonstrate-Understanding* from the group of categories collectively called *Rote-And-Routine* as used in the PET.

Despite the fact that rote learning "may be the most disparaging phrase in the educationists' glossary" (Hirsch, 1996, p. 59), the meaning of 'rote should be articulated for common interpretation. Ausubel (1962), Ivie (1998), and Moreno and Tharp (2006) distinguished rote learning from meaningful learning. Rote learning has been described as the learning of discrete elements of information in an arbitrary way without knowledge of the connection between the separate elements (Ausubel, 1962). Rote learning has also been described as the inability to transfer of knowledge to a new situation, because the knowledge was not understood (Mayer 2002). Understanding implies that the information has been placed into the student's existing, organized system of knowledge (Ausubel, 1968). If meaningful-learning occurs "when we grasp the relationship between two or more ideas, old and new" (Ivie, 1998, p. 39) and thought-learning is the ability to use memorized facts (Miller et al., 1978; Willingham, 2007), then "meaningful-learning" and "thought-learning" and "thinking" could be used synonymously with "understanding" as used in the PET. However, the term 'rote' is used somewhat differently in the PET, because it does not assume that connections have necessarily been made between the bits of information solicited by a question. Instead it is used rather to signify that the learner is able to recognize or recall the bits of information. The use of 'rote' in this way by the PET is similar to the Mayer (2002) description of rote learning: as being concerned with the retention of knowledge associated with cognitive processes of recognizing and recalling.

The PET category labeled as Perform-Routine-Procedures is included in the PET mega-category called Rote-And-Routine because the performance of procedures occurs in a routine way. That is, the performance of routine procedures show that a student exhibits some of the routine skills associated with science, but the student is not required to show any evidence of understanding the performance. Bloom et al. (1956, p. 201) included "recall of methods and processes" in their Knowledge category which means the PET mega-category called Rote-and-routine is synonymous with remember of the BTEO.

Irrespective of which theory of understanding is invoked, different topics, different disciplines and the experience and development of people will complicate how we view understanding (Perkins, 1998b). Understanding has been viewed as "the ability to think and act *flexibly* with what one knows" (Perkins, 1998a, p. 40, emphasis added), the ability "to go beyond knowledge and routine skill" (Perkins, 1998a, p.42), the ability to see connections between what students learn in school and their everyday lives (Perrone, 1998), and working with cognitive levels beyond knowledge of facts (Haladyna, 1997; Miller et al.,1978). Understanding can therefore be considered to be dependent on prior knowledge and on the context in which the prior knowledge can be used (Willingham, 2007). The groupings of performance expectations into a mega-category called Demonstrate-Understanding in the PET reflects all these notions of the meaning of understanding and acknowledges the existence of degrees of understanding, as described by Perkins (1998a). Understanding can denote deep coverage, or simply depth, and deeper understanding can signify better knowledge integration (Schwartz et al., 2009).

The PET defines Explain as the lowest level of understanding by which students are required to show evidence that they understand only what is explicated in the CBS. The PET categories Analyze and Apply denote deeper levels of understanding because these categories predicate a basic understanding of the knowledge and routine skills explicated in the CBS as necessary for the construction of new knowledge, or for the solving of new problems, beyond what was stated in the CBS. This view of deeper understanding is consistent with that of Mansilla and Gardner (1998) who believe that understanding is multidimensional and that deep understanding requires that students can use knowledge simultaneously in a number of different dimensions. Ormell (1974, 1979) considered the original BTEO to be lacking because it included "understanding"[123] as a constituent part of the taxonomy rather than as a collective descriptive term which might have been better used to incorporate some or all of the other categories. However, in a reference to a description of the BTEO categories beyond Comprehension, Bloom et al. (1956, p. 2) wrote that "there are other levels of the taxonomy which the teacher could use to indicate still deeper 'understanding' " which implies that they recognized levels of understanding, and that understanding was more than what their category called Comprehension embraced. Bloom et al. (1971, p. 272) confirmed that the BTEO implied levels of understanding when they described the BTEO category Comprehension as representing "the lowest level of understanding". Demonstrate-Understanding as used in the PET as a mega-category, is synonymous with what Bloom et al. (1956) in the BTEO call "intellectual abilities and skills" (of which Comprehension is one part).

Anderson et al. (2001, p. 70) noted that students are able to understand "when they are able to construct meaning from instructional messages … however they are represented" and "when they are able build connections between the 'new' knowledge to be gained and their prior knowledge". Similarly, students demonstrated an understanding of biology when they could apply knowledge to new tasks (Schönborn & Bögeholz (2009). The PET mega-category Demonstrate-Understanding conceptualizes understanding slightly differently from that of Anderson et al. (2001) and Schönborn and Bögeholz (2009) because neither the PET Explain nor the PET Analyze relate to "new" knowledge but knowledge that is stipulated to be learned in the context of the CBS. In the PET, the "new" knowledge refers to implicit connections between the content and the skills explicated in the CBS (PET categories C and D) and to unfamiliar contexts (PET category E). The PET mega-category Demonstrate-Understanding does embrace the "construct meaning" of the Anderson et al. (2001, p. 70) definition, as Explain requires a student to demonstrate their meaning of the knowledge and routine skills required by the CBS. Basic understanding (PET Explain) conceptualized in the PET, as a link between Memorize and Perform-Routine-Procedures, and the deeper levels of understanding (PET Analyze and Apply). This view is consistent with views from cognitive science that state that "[s]tudents need to learn [PET category C] the core-concept framework [PET categories A and B] that underlies instruction because it facilitates their cumulative development of in-depth understanding [PET categories D and E]" (Vitale, Romance, & Dolan, 2006, p. 3).

123 Many authors, including Bloom et al. (1956), used "understanding" as synonymous with "Comprehension" when talking about the BTEO.

What is memorized or performed routinely is of little or no consequence by itself (Gardner, 1998; Perkins, 1998a; Miller et al., 1978), but relevant content knowledge is necessary for students to be able to make the judgments associated with understanding (Senechal, 2010). According to the relationships that have been articulated between the five PET categories, understanding is *dependent on* the knowing of facts and concepts and the ability to be able to perform specific routine procedures, the rote and routine. This kind of relationship envisaged by the PET recognizes the importance of factual knowledge and routine activities in developing student's academic competencies (Booker, 2007; Hirsch, 1996; Willingham, 2007) and the understanding necessary for constructing new knowledge, framing new problems and solving them (Mayer, 2002; Roth, 1993; Yager, 1993). The positioning of Apply in the PET is consistent with the work of Miller et al. (1978) who collectively labeled the cognitive levels they considered higher than factual recall (i.e., Summarizing, Predicting, Evaluating and Applying), and inferred that the most comprehensive of levels of thinking was applying what had been learnt.

b. ***Lower Order Cognitive Skills vs Higher Order Cognitive Skills***

A motivation for the recognition of the second grouping of the five PET categories into two mega-categories, LOCS and HOCS[124] came from the CBS. The CBS stated that for HG "[a]pproximately 60% of the marks allocated to the questions in the written examination will be recorded for recall of knowledge. The remaining 40% will be awarded for higher skills" (DNE, 1984a, p. ii).[125] However, because the CBS gave no direction as to how HOCS could be identified, teachers and examiners defined the skills in different ways (Figure 4.3), and this has implications for equivalence when comparing standards (Section 4.1). What follows is an attempt to construct a meaning for the way that the PET recognizes HOCS and LOCS based on how different authors have viewed these categories, and their relationship with one another in the literature.

Hancock (1994) and Leighton (2011) both lamented that the lack of definition and description of those cognitive skills subsumed by the term *higher level thinking* in the literature made it difficult to interpret the conclusions reached by many studies. At best, the concept of HOCS has "served to remind us there is more to learning than the mere memorization of facts and figures … [and] to teach higher order thinking skills one needs criteria for making such judgments [as to what HOCS are]"(Ennis, cited by Ivie 1998, p. 35). To illustrate the need for

124 The author recognizes that the PET categories are only a proxy for cognitive demand categories but has included the term 'cognitive' here to avoid introducing new acronyms into the discussion. LOCS and HOCS were used in Sections 4.1 and 4.2.

125 For SG the proportions were: 75% HOCS and 25% LOCS (DNE, 1984b, p. ii).

an operational definition for HOCS, Champagne (1990) used the example of two science programs in the USA each of which approached teaching in different ways, based on perceptions of the relationship between LOCS and HOCS. In one program, HOCS were considered a concatenation of LOCS, and thus when LOCS were taught, HOCS were assumed to be learned. In the other program, HOCS were considered to be different from concatenations of LOCS, and HOCS were considered to develop from the understanding of phenomena (Champagne, 1990). Domin (1999, p. 109) argued that “higher order cognition presupposes lower-order cognition” and used the cumulative hierarchical nature of BTEO to support his position.

A number of authors and the CBS have incorrectly implied that Bloom et al. (1956) defined which of the six BTEO categories referred to HOCS. Bloom et al. (1956, p..38) collectively called ‘critical thinking’, ‘reflective thinking’, and ‘problem solving’ “intellectual abilities and skills”, and recognized within this category increasing higher levels of problem-solving as one moved up the BTEO hierarchy of categories. Domin (1999) and Ivie (1998) claimed that Bloom's ‘top’ three levels (i.e., Analysis, Synthesis, and Evaluation) constituted the HOCS. Other authors, like Zohar and Dori (2003) both included and excluded the BTEO category Application when talking about HOCS in the same paper. Miri, David and Uri (2007, p. 355) claimed that higher order thinking “corresponds” with the BTEO “overlapping levels above [C]omprehension. Accordingly, recall of information would be an example of a lower order cognitive pattern, or thinking skills, whereas [A]nalysis, [E]valuation, and [S]ynthesis would be considered higher order thinking skills”. Thus, Miri et al. (2007) include the BTEO category Application in their one definition of HOCS and exclude in their other definition of HOCS. Section 4.2.3 above documented further confusion in how HOCS are viewed in a set of research papers associated with two recent papers, (Haak et al., 2011; Freeman et al., 2011) Similarly, different interpretations of HOCS within and between different interpretations of the CBS, either included or did not include the BTEO category Application with the HOCS (Section 4.1).

Zoller and Tsaparlis (1997) described LOCS questions as requiring the simple recall of information or a simple application of procedures, which were not necessarily understood, to familiar situations and contexts, in contrast to HOCS questions requiring analysis, synthesis, problem solving capabilities, the making of connections or applying critical evaluative thinking to unfamiliar situations. The categories of PET considered as LOCS all function within the familiarity of the knowledge and the skills explicitly stipulated in the CBS, and HOCS includes those categories in which the context is not stipulated in the CBS. Other authors (e.g., Bransford, Brown & Cocking, 1999; Champagne, 1990; Knapp, 1992; Marzano

& Kendall, 2007; Perkins, 1993a, 1993b; Perkins & Salomon; 1989) have drawn on work in cognitive psychology to demonstrate that the use of HOCS is closely linked to discipline-specific knowledge. Hirsch (1996) described higher-order thinking as having a mixed character requiring both breadth of factual domain-specific knowledge, and operational facilities in the form of the domain-appropriate procedures and strategies for dealing with the facts. Similarly, when Silva (2009, p. 630) described "[a]n emphasis on what students can do with knowledge, rather than what units of knowledge they have" as "the essence of 21-st century skills", she distinguished between HOCS and LOCS. She also reiterated that what was important for teaching and learning was not either LOCS or HOCS, but both. Accordingly, HOCS—as conceptualized in the PET—indicates that the HOCS categories Analyze and Apply draw on a basic understanding the facts, concepts and routine procedures captured collectively as LOCS. If science concepts have been learned this is evidenced most strongly when students can use concepts in real life situations (National Science Teachers Association, 1982) and the ability to select relevant learned knowledge, and to apply it to new novel situations, is considered to be necessary for scientific literacy (Enger & Yager, 2001). Thus, the conceptualization of HOCS in the PET recognizes that "[p]rior knowledge is the beginning of new knowledge. It is always where learners start" (Zull, 2002, p. 93).

4.3.2 Validation of the Performance Expectations Taxonomy

A list of 37 possible performance expectations with which high school Biology students might be expected to engage, was compiled from discussions with teachers, policy documents and the literature, for example, CCSSO and WCER (2004). The PET was developed in this study to, first ensure that meaningful and unambiguous categories of performance expectations could be recognized to summarize, describe and compare the cognitive demand of the different SC Biology examination question papers analyzed in this study. Second, because classification of the performance expectations of questions is based on human judgement, inter- and intra-judge consistency was important for the trustworthiness and consistency of the categorization of cognitive demand using the PET. Given that 11 006 scorable events[126] had to be analyzed by the author in this study meant that multiple raters were not available to classify the performance expectations for all of the scorable events. Therefore, it was necessary to ensure that there would be a high level of consistency between the author's classifications using the PET and other judges who used the PET to classify the same performance expectations of a given subset of questions. The claims that the PET could perform these two functions had to be validated before the PET could be used for the broader study.

[126] Scorable events are questions or sub-questions which cannot be further broken down for analysis. They are described more fully in Chapter 5.

Validation of the PET instrument required the purposeful selection (McMillan & Schumacher, 2001) of experts in the field of secondary school biology education. While defining an expert in science education is not an easy task (Osborne, Collins, Ratcliffe, Millar & Duschl, 2003), Häussler and Hoffmann (2000) recommended that experts be individuals who are reflective of their professional field, are open to the opinions of other people, and are actively involved in promoting teaching and learning. Four such experts were identified and all agreed to participate in the validation process. All four persons had been involved in both educational research programs and in teacher development, and all had extensive expertise in the teaching of Biology at Grade 12 level in South African schools. One was a Professor of Science Education at a university, one was a retired lecturer of pre-service student Biology teachers at a teachers' training college and at a university, another was a Senior Curriculum Planner for Life Sciences at one of the provincial education departments, and the fourth was an experienced former high school Biology teacher who was involved in teacher education while completing her PhD in science education. Two of the four participants were involved with the setting of, or the moderation of, the national SC Biology examinations.

The four participants were given an overview of the PET and the philosophy which guided the development of the PET (Section 4.2). Additional documents provided to the participants, and discussed by the participants and the author, were a) instructions to the participants about rating and coding the performance expectations of a question paper (Appendix 4.4); b) a table of possible topics (Appendix 4.5);[127] c) a table of possible performance expectations (Appendix 4.2); d) a dichotomous classificatory key to determine the category of PET for the performance expectation identified in a question (Appendix 4.3); e) examples of the classifications of questions using PET (Appendix 4.6); f) a copy of the CBS and the modified CBS and g) data capture sheets for the particular examination. Once the author had explained these documents and the rationale behind the PET, copies of the 2005 HG examination question papers, (i.e., Paper 1 and Paper 2), and their corresponding marking memoranda (DoE, 2005a, 2005b), were given to the participants (known from here on as 'raters').

Each rater was required to independently classify each of the performance expectations solicited by each of the sub-questions in the two examination question papers. Raters were asked to code the performance expectation of each identified question at only the level of category, that is: Memorize; Perform-Routine-Procedures; Explain; Analyze or Apply (A to E)

[127] Participants were not required to classify questions according to topic. The table of possible topics was given so that participants could locate appropriate data in the CBS when classifying the performance expectations.

(Appendix 4.2 and Appendix 4.3). The descriptors listed for each category (Appendix 4.2) were to be considered to be illustrative of the types of activities (performance expectations) associated with each category and were not to be used for coding purposes.

Once all the raters had completed the classification of all the questions, the author put the raters' classifications together with her own classifications in a table for each question paper (Appendix 4.7 and Appendix 4.8) for discussion by the author and the raters. Each classification made by the group was discussed in detail, and a performance expectation was negotiated for each classification, and helped to focus the descriptions of the PET categories.

The negotiated classification was then added to each of the tables (Appendix 4.7 and Appendix 4.8). The inter-rater reliability concerning the assigning of the PET categories to determine cognitive demand was analyzed by determining the percent agreement between each of the five sets of ratings (author plus four experts) and the classifications negotiated by discussion between the combined group of five raters. A summary of the inter-rater agreement with the negotiated agreement for HG Paper 1 is shown in Table 4.1. Two points arose during the group discussion about the 2005 HG Paper 1 performance expectation classifications from which the negotiated performance expectation classifications emerged:

Table 4.1 Percentage agreement between the PET classifications of individual raters and negotiated PET classifications of the group of five raters.

	Rater				
Question Paper	**AC**[a]	**1**	**2**	**3**	**4**
Paper 1	87.5 [b]	60.0	58.8	68.8 [b]	40.0
Paper 2	93.6 [b]	50.0	48.7	60.2 [b]	35.0
Paper 2 [c]	— [d]	98.7	98.7	98.7	97.4

Note:
a Author of this thesis.
b Rater used the classificatory key.
c All raters used the classificatory key.
d Author's classification was used for comparison.

1. There was a tendency for raters to confuse the complexity of the instruction of a question with the performance expectation or the science of the question. For example, raters agreed that a multiple choice question which required a student to recognize a function of the human large intestine (Figure 4.9 [A]) should be categorized as a Memorize question (PET category Explain). Another multiple choice question which

required a student to recall the pH of three digestive juices and then to arrange the pH values in order of smallest to largest (Figure 4.9[B]) was considered by some raters to be a category Explain or a category Analyze question because of the complexity of the question. On discussion, the group consensus was that the second question was asking the student to do no more that memorize—the complexity resulted because of what the student was required to do with the recalled pH values.

National 2005 HG Paper 1

A.

1.1.1 What is the main function of the large intestine in the human body?

A Digestion of cellulose
B Absorption of water
C Storage of nutrients
D Absorption of digested food

B.

1.1.2 Arrange the following digestive juices in the order of their pH values from the smallest to the largest.

(i) bile
(ii) gastric juice
(iii) saliva

A (i), (ii), (iii)
B (ii), (iii), (i)
C (ii), (i), (iii)
D (iii), (ii), (i)

Figure 4.9 An example of two questions classified according category Memorize of the PET, showing different levels of complexity within each question.

2. The best consensus occurred when raters used the classificatory key (Appendix 4.3) to classify the performance expectations. Raters 1, 2 and 4 used the table of PET categories (Appendix 4.2) rather than the classificatory key when they made their independent classifications. These raters indicated that the reason they did not use the classificatory key was because they were distracted by the illustrative types of activities (individual performance expectations) associated with each category given in the table of PET categories.

Due to time constraints, 2005 Paper 2 could not be discussed face-to-face. The author, taking cognizance of the discussion that flowed from the consensus building dialog around the Paper 1 PET classifications, compiled detailed explanations for what she thought, post-validation, the performance expectation classifications for Paper 2 should be. Each participant was sent the detailed classifications and explanations (Appendix 4.9) and was asked to compare them to what their own, post-validation, classifications would be for the same questions if they used the classification key. The results of the classifications of Paper 2 questions using the classificatory key are shown in Table 4.1. As can be seen, using the classificatory key resulted in much improved inter-rater agreement. In all cases where a rater disagreed with the author's classification, the discrepancy was the result of different interpretations of the syllabus and, after discussion with the author, each rater ultimately accepted the author's classifications (Table 4.1). The raters examined, and agreed with all of the author's PET classifications for the 2006 HG Paper 1 and Paper 2.

The resulting consensus that was reached between the expert raters themselves and between the expert raters and the author during the validation and post-validation exercises determined that subsequent classifications of the performance expectations that the author had to make during the analysis of examination question papers, would most likely be trustworthy and consistent. As a result of the dialog between the author and the four expert raters, minor changes in wording were made to both the classification table which describes each category of the PET and gives examples of the types of activities associated with each of categories, and the classificatory key to ensure that each category is unambiguous, and thus improve the meaningfulness of the PET

4.3.3 <u>The Performance Expectations Taxonomy, Bloom's Taxonomy of Educational Objectives and modifications of Bloom's Taxonomy of Educational Objectives</u>

This section discusses the conceptual relationship between the PET, which is used in this study to determine cognitive demand, and BTEO the taxonomy of cognitive demand used in the SC Biology policy documents and (Section 4.3.3.1). This conceptual relationship is described by referring to two other similar modifications of BTEO that were made to describe cognitive demand in studies of Biology teaching, learning and assessment, that is, Crowe et al. (2008) and Zohar et al. (1998). Practically, to test the described conceptual relationship between PET and BTEO, the actual classifications of a subset of SC Biology examination questions made by examiners using BTEO are compared with classifications of the same questions made by the author using the PET (Section 4.3.3.2).

4.3.3.1 *Conceptual relationships between the taxonomies*

This section describes some similarities and differences in the way that categories of cognitive demand are viewed by the PET, BTEO and two modifications of BTEO (Figure 4.10). The two modifications of BTEO are included here because both studies resulted from the use of BTEO in studies of Biology teaching and learning─the same discipline in which this study is rooted─and both studies conceptualize the BTEO category Application in both similar and different ways to which this author conceptualizes the category Apply in the PET (Figure 4.10). The arguments of Zohar et al. (1998) and Crowe et al. (2008) in their modifications of BTEO—in particular how they conceptualize the BTEO category Application are used to support the structure of the PET. The focus here will be on both the position of BTEO Application category and the way Apply is conceptualized is in the PET, and the PET category, Perform-Routine-Procedures which was not recognized by either BTEO or the two adaptations of BTEO mentioned.

One obvious difference between BTEO and the PET is the terminology used to describe the first of the categories, that is, *Knowledge* in BTEO and *Memorize* in the PET. However, descriptions of what these categories represent within each taxonomy are similar in many respects. They differ in how Knowledge was described in the BTEO[128]and how the concept of knowledge has been incorporated in the PET. One such difference is that BTEO description of Knowledge includes the knowledge of methodology which the PET includes in the separate category Perform-Routine-Procedures.

The BTEO category Application has the shortest description of all the categories recognized by the taxonomy. The authors of BTEO described Application as constituting a problem which is "new to the student, [but] he will apply the appropriate abstraction without having to be prompted as to which abstraction is correct or without having to be shown how to use it in that situation". Moreover, how a student proceeds with the problem "would depend on the student's familiarity with the problem" (Bloom et al. 1956, p. 120). While Bloom et al. (1956) explained the terms "unfamiliar" and "familiar", their meanings are hard to interpret in practice, without a clearly defined context in which students are functioning. For this reason in South Africa the BTEO category Application is considered a LOCS in one interpretation of the CBS, and a HOCS in another (Section 4.1). The hierarchical nature of the BTEO implies that the category Application is subsumed into the higher levels of the taxonomy.

128 The interested reader is directed to Anderson et al. (2001) where the use of knowledge as a category in BTEO is discussed in an explanation for why a separate dimension for knowledge in a revision of BTEO. Marzano (2000) and Marzano and Kendall (2007) also discuss the BTEO use of knowledge in the light of current understanding of what is meant by knowledge.

A. Bloom's Taxonomy (Bloom et al., 1956)

Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
KNOWLEDGE	INTELLECTUAL ABILITIES AND SKILLS				
Lower order skills	→				*Higher order skills*

B. Modified Bloom's Taxonomy (Zohar et al., 1998)

Knowledge	Comprehension	Low Level Application	High Level Application [a]	Analysis	Synthesis	Evaluation
Lower-order thinking [a]			*Higher-order thinking*			

Note:
a Not named but implied from text

C. Modified Bloom's Taxonomy (Crowe et al., 2008)

Knowledge	Comprehension	Application	Application	Analysis	Synthesis	Evaluation
Lower-order cognitive skills			*Higher-order cognitive skills*			

D. Performance Expectations Taxonomy

Memorize	Perform-Routine-Procedures	Explain	Analyze	Apply
Lower-order skills			*Higher-order skills*	

Figure 4.10 Relationship between the Performance Expectations Taxonomy developed in this study, the original BTEO (Bloom et al., 1956) and adaptations of Bloom's Taxonomy used in Biology studies (Crowe et al., 2008; Zohar et al., 1998).

The PET uses Apply to refer to contexts that are explicitly outside of the boundaries of the CBS, and are therefore unfamiliar. In the case of the PET it is therefore appropriate to place Apply at the ‘top’ of the taxonomy. Since BTEO Application does not include creative problem-solving, it could not be positioned at the apex of BTEO. Given views that many of BTEO higher order categories merge into one another (Lord & Baviskar, 2007; Nordvall & Braxton, 1996) and the author’s own experience using BTEO, the PET Apply category includes BTEO categories, Synthesis and Evaluation, as well as the unfamiliar aspects of the BTEO category Application.

In another description of the BTEO category Analysis, the authors write that “no entirely clear lines can be drawn between [A]nalysis and [C]omprehension at one end or between Analysis and Evaluation at the other” (Bloom et al. 1956, p. 144). This claim ignores the feature of the BTEO construction, whereby getting from Comprehension to Analysis requires invoking Application. “Such an ability [analysis] may be regarded as a further “comprehension” of an idea, problem or document” (Bloom et al., 1971, p. 177) again suggests an omission of the BTEO category, Application from conceptualizations of the BTEO. In an analysis of the relationships between the categories of the BTEO using student performance in science, Madaus, Woods and Nuttall (1973) noted a strong indirect link between the non-adjacent levels Comprehension and Analysis. Similarly, in her study to test a curriculum about teaching fifth and sixth graders system dynamics thinking, Roberts (1976) experienced difficulties in the positioning Application type questions within the BTEO. Concerned that the directions to the Application questions used in her test might have been unclear, they were rewritten prior to testing, and yet students still had questions about these item when the test was adminstered. In another study, Klein (1972) found that all levels of BTEO except for Application could be elicited and a range of performance detected among seven- to nine-year-olds in social studies tests. Stedman (1973) also questioned the relationship conceptualized in the BTEO, between Comprehension and Application

The above discussion suggests that the category Application as described in BTEO is problematic when used in practice. Two studies, both concerned with Biology teaching and learning, used modifications of BTEO (Zohar et al., 1998; Crowe et al., 2008). Both of these studies recognized that BTEO Application could encompass two levels, that is, a lower level and a higher level (Figure 4.10). Scrutiny of the assessment questions considered by both these studies[129] indicates that these two levels of Application questions are accommodated in different ways. Questions classified as lower-level Application by Zohar et al. (1998) would

[129] Specimen questions (A[lison]. Crowe, personal communication, July 9, 2010; Zohar et al.,1998).

have been classified in the PET category Perform-Routine-Procedures because they required students to do things routinely required with the study of biology, which is consistent with the view that these actions are not HOCS. Unfortunately, Zohar et al. (1998) gave no examples of questions that they would have considered as higher-level Application. Crowe et al. (2008) coded some Application questions as both LOCS and HOCS, and 'advanced' Application as HOCS only.

The PET separates actions that might have been considered BTEO Application into an additional category Perform-Routine-Procedures and into Apply. The placement of Perform-Routine-Procedures in the PET recognizes that there are routine performances which are distinguishable from the facts, definitions, concepts, and formulae, which are fundamental to the learning and understanding of science. These routine performances are no more than LOCS if they are considered to be within the explicit boundaries of the CBS. The positioning of Apply in the PET recognizes the importance of the 'unfamiliar' in the way that Apply is conceptualized. The implications of the different ways that Application was used in BTEO and Apply is used in the PET are discussed below.

The PET described above was developed to determine the cognitive demand of S C Biology examination questions, and could in principle apply to interim or formative assessment in any subject. A change of wording (i.e., redefining familiar and unfamiliar to the relevant context) in the PET would make this instrument suitable to use for classifying the cognitive demand of tasks used in summative assessment tasks in Biology and other subjects. This is, however, beyond the scope of this thesis.

4.3.3.2 *Calibration of the taxonomies*

Each method of classification, because it conceptualizes cognitive demand in a particular way, tells its own story and therefore different methods should not be used interchangeably when making inferences about the cognitive demand of examinations. However, if two different methods of classifying questions with respect to cognitive demand can be calibrated with one another by classifying the same questions by both methods and comparing the resulting classifications, and operational relationships can be established between the two methods. For example, inferences made from the use of one of the methods can be made relative to the other method. As it was argued above (Section 4.3.3.1) that the PET is a modification of BTEO and both the national DoE and the IEB examination policies required that SC Biology examinations were set using BTEO (Section 4.1), it was necessary to try to calibrate BTEO and PET in two different ways. The first way involved classifying each of the performance expectations listed in the PET (Figure 4.6) using the BTEO, and the second involved re-classifying using the PET a set of examination questions that had previously been classified by the examiners[130] who set SC Biology papers using the BTEO.

The classification of the performance expectations listed in the PET using BTEO are shown in Figure 4.11 and will not be discussed here because this figure demonstrates the conceptual relationship between the PET and BTEO already explained in Section 4.3.3.1 above. Instead, what follows focuses on the calibration of the two taxonomies using examination question to elucidate the relationships—in practice—between the two taxonomies.

A request was made to the national DoE for the records of classifications of the cognitive demand made by the panel of examiners who set the national Biology SC question papers for 2005 and 2006 using BTEO.[131] Unfortunately, records of cognitive demand could only be located for two, that is, HG and SG 2006 Paper 2, of the eight question papers requested.[132]

Summaries of the classifications of the questions in the two papers, 2006 HG Paper 2 and 2006 SG Paper 2, made by the DoE using the BTEO, and those made by the author using the PET are shown in Figure 4.12. As expected, because the PET claims to be a modification of BTEO (Section 4.3.3.1), there are similarities in the way that questions were classified using each of the two taxonomies for both question papers. However, it is the differences between the classifications made with each of the taxonomies that are the most interesting here, because they demonstrate the usefulness of the PET for the broader study encompassed by this thesis.

One obvious similarity that emerged, because of how these categories are defined within the respective taxonomies, was that all questions classified as BTEO category Knowledge questions were classified as the PET category Memorize questions for both the HG and the SG papers. Questions classified in by the examiners as BTEO category Comprehension gave a different story—some were classified in each of the PET categories, Memorize, Perform-Routine-Procedures and Explain (both HG and SG question papers), and Analyze and Apply (HG question paper only). The classification of some of the BTEO category Comprehension questions showed evidence that the examiners had conflated the complexity of the instruction of a question with the knowledge of science required by the question and therefore,

[130] A panel of four examiners and one internal moderator reached consensus about the classification of questions using BTEO.

[131] In this study candidates answer scripts are analyzed for the years 2005 and 2006.

[132] Requests were made via Dr R Poliah (Director: Examinations and Assessment) and Mrs P Ogunbanjo (Deputy Director: Examinations and Assessment) of the national DoE.

Knowledge ī recognition or recall of specifics, universals, methods, processes, patterns, structures or settings *1.10 Knowledge of specifics* *1.11 Knowledge of terminology* *1.12 Knowledge of specific facts* *1.20 Knowledge of ways and means of dealing with specifics* *1.21 Knowledge of conventions* *1.22 Knowledge of classifications and categories* *1.24 Knowledge of criteria* *1.25 Knowledge of methodology* *1.30 Knowledge of universals and abstractions in a field* *1.31 Knowledge of principles and generalizations* *1.31 Knowledge of theories and structures*	**Comprehension** ī lowest level of understanding of given or remembered knowledge *2.1 Translation* *2.2 Interpretation* *2.3 Extrapolation*	**Application** ī use of knowledge in particular and concrete situations	**Analysis** ī the breakdown of knowledge into constituent elements to reveal relationships and organizational principles *4.1 Analysis of elements* *4.2 Analysis of relationships* *4.3 Analysis of organizational principles*	**Synthesis** ī to re-arrange or combine knowledge to convey ónewô knowledge *5.1 Production of unique communication* *5.2 Production of a plan, or proposed set of operations* *5.3 Derivation of a set of abstract relations*	**Evaluation** ī make judgment about knowledge *6.1 Evaluation in terms of internal evidence* *6.2 Judgments in terms of external criteria*
1, 2	15	3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 27	16, 17, 18, 19, 20, 21, 27, 23, 24	17, 18, 19, 20, 21, 22, 23, 24, 29, 30, 31, 32, 33, 34, 35, 36, 37	20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37

Note:
a The numbers 1 to 37 refer to the performance expectation numbers used in the PET (Figure 4.6)

Figure 4.11 The PET cross-referenced with BTEO (Bloom et al., 1956).

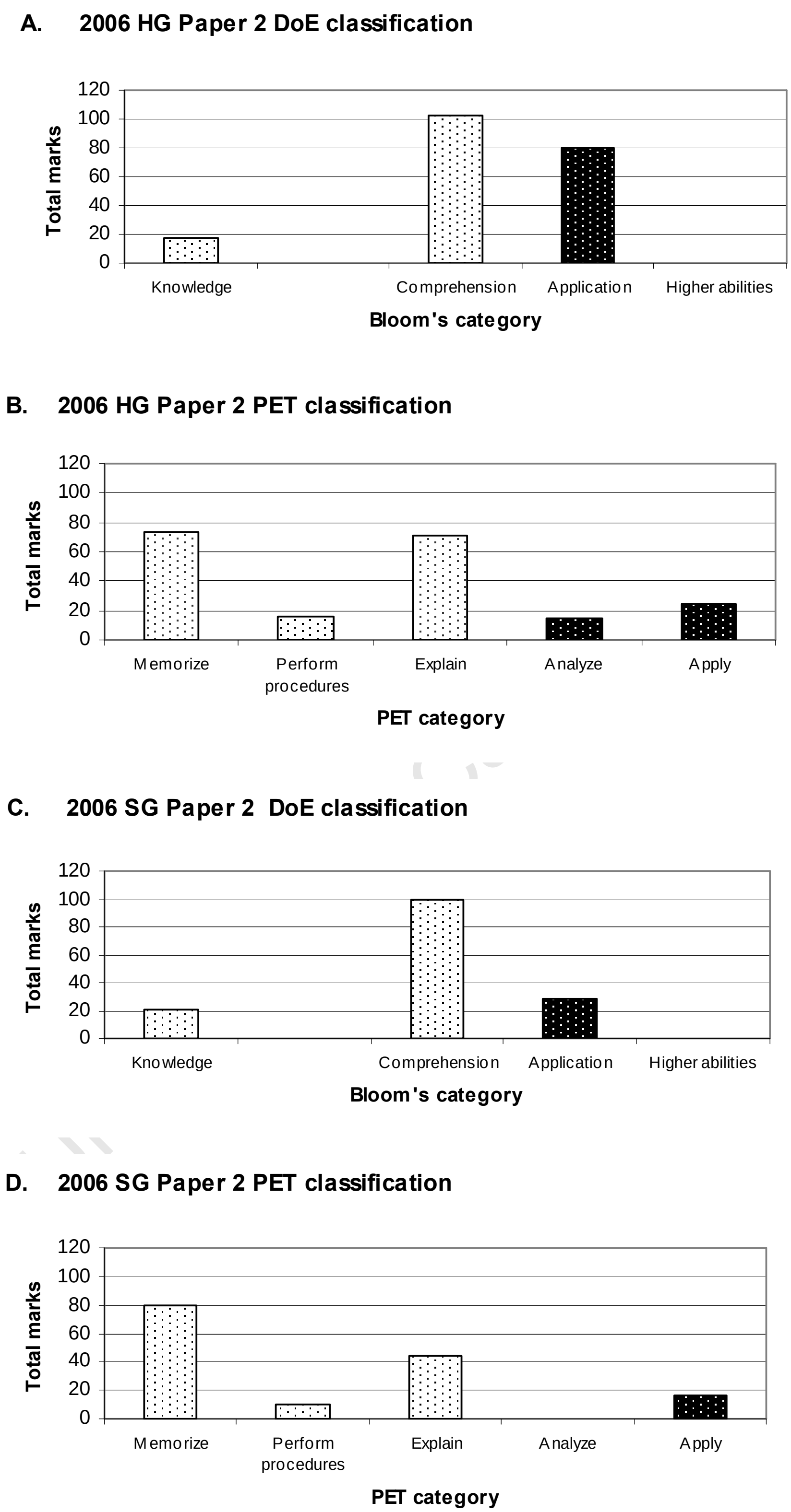


Figure 4.12 Comparison of DoE classifications and PET classifications of cognitive demand. Dark shading indicates the HOCS recognized by the DoE interpretation of BTEO and the PET.

classified some questions as higher than BTEO Knowledge. For example, a multiple choice question which required students to select the correct combination of given functions of the human eye, were classified by the examiners as a BTEO Comprehension question. Similar questions were discussed during the validation of the PET and a consensus decision was made that such questions be classified as no more that the PET category Memorize (Figure 4.9). Other questions classified as BTEO Comprehension involved the identification of labelled parts of the human eye in a drawing, and recalling the function of another labelled part of the same drawing. As the syllabus required that students know the structure and function of different parts of the eye, these questions were classified in the PET category Memorize.

Questions classified in BTEO category Application included some questions which were classified in each of the PET categories: Memorize, Perform-Routine-Procedures, Explain and Apply (both HG and SG question papers) and Analyze (HG question paper only). The one question classified as BTEO Application and the PET Memorize, concerned the identification of the highest concentration of auxins in two different positions marked on a diagram of a germinating seed. Answering this question would not convey any understanding of growth in plants because students were required to learn about the effect of gravity on hormone distribution in plants. Another question classified in BTEO category Application and the PET Perform-Routine-Procedures involved calculations and the drawing of a biological structure and a graph, routine skills which indicated no specific understanding of the biological meaning of the calculations or of the data represented by the graph or by the drawing. Questions which were classified as BTEO category Application and the PET category Apply involved a new context as defined by both these categories in their respective taxonomies.

Neither the HG nor the SG question papers had questions classified in BTEO categories Analysis, Synthesis or Evaluation, which the DoE collectively groups in a category called ‘higher abilities’ (Section 4.1). The implications of this are profound in terms of the assumed equivalence between the national DoE and the IEB question papers. If the DoE question papers were interpreted using IEB policy (Section 4.1) the DoE question papers would have had no questions, classified using BTEO, which addressed higher order cognitive skills. The implications of these differences in policy were discussed in Section 4.1.

Classifications according to the PET resulted in some questions from the BTEO categories Comprehension and Application being grouped in the PET category Analyze. Some of these questions required students to access content detailed in two different topics given in the syllabus (e.g., nervous and chemical co-ordination and thermoregulation). When a question requires students to connect explicated content knowledge from two different areas of the syllabus (familiar context) to answer the question, the question should be classified as the PET category Analyze. Two other questions classified as the BTEO category Comprehension and the PET category Analyze required the students to connect information from a graph showing changes in human skin and core temperatures over a period of time with what they had learned about human temperature (i.e., make deductions in a familiar context).

Figure 4.13 describes how an essay-type question was classified differently by the PET and by the examiners using BTEO. The national DoE marking memorandum (Figure 4.13 [A]) awarded “any 15 [marks]” for the main part of their essay. Students did not need to give the answer in the correct sequence, but they could simply recall what they knew about water movement into a plant. part of the question as the BTEO Comprehension and the PET Memorize. The synthesis part of the question recognized that students should show a basic understanding of they were expected to have been taught and was classified as PET Explain, and the examiners classified it as BTEO Application. The marking memorandum for the IEB essay allocated essay marks in a way that recognized the importance of how facts and concepts are used in a structured way to answer the question (Figure 4.13 [B]. Accordingly the three different sections of the IEB essay question were classified using PET as (i) Memorize, because only accuracy of content was required—24 marks; (ii) Explain, because some understanding of how to put together the learned content was required—28 marks; and (iii) category Perform-Routine-Procedures, because only the presentation of the work was assessed—8 marks. This figure, Figure 4.13 also shows the importance of using the marking guidelines to clarify what examiners are looking for in a question, before classifying a question. The example IEB essay illustrates the importance of the wording of a question to convey to both the candidates and to the examiners explicit expectations about how a question should be answered and how it question will be marked.

From the above comparison of the BTEO and the PET, two advantages of using the PET emerge. First, the PET category Perform-Routine-Procedures highlights that not all questions requiring a demonstration of routines are application questions—many as asked in the SC Biology examinations are routine procedures— and cannot be considered as HOCS questions. For example, drawing a graph, does not require higher intellectual skills. Second, in the PET the familiar is clearly linked to the scope of the CBS, and the unfamiliar is linked to contexts outside of the CBS.

A. National Department of Education 2006 HG Paper 2 (compulsory essay)

5.2 The root hair is structurally suited for its function. Explain this statement in the light of the process by which the root hair absorbs water from the soil. Describe how the absorbed water is then transported to the xylem of the root.

Content: (15)
Synthesis: (3)

Marking memorandum:

Content:
lists 19 points of which a maximum of 15 were to be marked

ī [t]he soil water has a higher water potential √
ī [t]he root hair has a lower water potential √
Ź
ī and into the xylem of the root √ (any 15)

Synthesis:

Level descriptions	Marks
Did not attempt the question.	0
Poor structuring of the answer with significant gaps in the knowledge of concepts and the adaptations of the tissues.	1
Answer is structured in a superficial way, illustrating the passage of water, [showing] some gaps of knowledge of concepts. The adaptations of the various tissues have been included as an add on (separated from description of the passage of water)[.]	2
Answer is well structured, and logically describes the sequence of the passage of water, together with the associated concepts. The adaptations of the tissues are explained/embedded in this logical sequential description.	3

B. Independent Examinations Board 2000 HG

Section C (choice of two essays)

Instruction: Your essay will be marked for:

(i) Accuracy relevance and completeness of content (24)
(ii) The skill with which this factual content is used to answer the question (28)
(iii) The organization of your work and the neatness and legibility of your presentation (8)

6. The cellular processes of photosynthesis and respiration are processes which complement each other to ensure that individual organisms survive, yet these processes also keep ecosystems functioning.

Discuss the importance of these TWO processes at these TWO different levels. (60)

Figure 4.13 Examples of PET classification of essay questions.

4.4 The relationship between the objectives and approaches of the Senior Certificate Core Biology Syllabus and the Performance Expectations Taxonomy

The CBS (DNE, 1984a, 1984b) provided a general list of learning objectives and the approach to the syllabus at the beginning of the document which were expected to apply **across all** the elaborated content SC students were expected to learn and on which they would be assessed (Chapter 2, Figure 2.4). Both the learning objectives and the approach to the syllabus inherently invoked cognitive skills,[133] and so it was necessary to cross reference both with the PET (Table 4.2 and Table 4.3) to explicate the relationships between PET and the aims of the CBS.

The attitudes and values which were listed in some of the learning objectives in the CBS are concerned with a respect, love, and an appreciation of nature, which cannot be directly measured by the PET (Table 4.3) nor could they be measured by the cognitive domain of the BTEO. The BTEO considered attitudes and values as a part of a separate domain, namely, the affective domain (Bloom et al., 1956). In writing their taxonomy of educational objectives for the affective domain, Krathwohl et al. (1964) recognized the possibility of some overlap between their conceptions of the cognitive and affective domains. The BTEO has "been justly criticized because it isolates aspects of the same objective—and nearly every cognitive objective has an affective component" (Anderson et al., 2001, p. 258). Anderson et al. believed that a recognition of metacognitive knowledge in their revision of BTEO "bridges the cognitive and affective domains" (Anderson et al., 2001, p. 259). Similarly, Anderson et al. (2001), Marzano (2001), Marzano and Kendall (2007), Wiggins and McTighe (1998, cited by Anderson et al., 2001), and Willingham (2007) argue that self-knowledge and metacognition make understanding, explanation, interpretation, application and critical thinking within a discipline more likely. While the PET does not explicitly invoke metacognition as a part of the taxonomy it acknowledges that metacognition is involved in the understanding of Biology at different levels (e.g., PET categories Explain, Analyze and Apply) and that understanding is necessary for students to start developing the affective attributes: a respect of, a love of and an appreciation of nature (Table 4.2). The PET categories Explain, Analyze and Apply, draw explicitly on the facts, concepts (PET Memorize) and procedures (PET Perform-Routine-Procedures) required by the syllabus. Therefore, the objectives of the CBS are included in all of the PET categories.

[133] Snow (1994) argued that in order to produce learning, affective and motivational processes should be viewed as interacting with cognitive processes, as in the case of the CBS.

Table 4.2 Objectives of the CBS used in this study (DNE, 1984a, 1984b) and their relationship to the categories of the PET.

Objectives	Performance Expectations Taxonomy category [a]
To develop in pupils the following important attributes:	
1. An understanding of fundamental biological principles based upon a study of living organisms;	C - Explain
2. An awareness of biological relationships;	A - Memorize B - Perform-Routine-Procedures C - Explain
3. An ability to make critical, accurate observations of biological material, and to make meaningful records of such observations;	B - Perform-Routine-Procedures C - Explain D - Analyze, or make connections between required knowledge or given information E - Apply concepts or procedures, or make connections, or make extended use of unfamiliar information
4. An ability to analyze and evaluate biological information, to formulate hypotheses and to suggest procedures to test them;	D - Analyze, or make connections between required knowledge or given information E - Apply concepts or procedures, or make connections, or make extended use of unfamiliar information
5. An ability to communicate clearly when reporting information and expressing ideas;	C - Explain D - Analyze, or make connections between required knowledge or given information E - Apply concepts or procedures, or make connections, or make extended use of unfamiliar information
6. A **respec**t for all living things and an urgent awareness of [hu]manôs responsibilities in the preservation of life, particularly in the S[outh]. A[frican]. context;	C - Explain D - Analyze, or make connections between required knowledge or given information E - Apply concepts or procedures, or make connections, or make extended use of unfamiliar information **respect** not measured
7. A **love** and **appreciation** for South African fauna and flora and a recognition of the urgent need for nature conservation.	C - Explain D - Analyze, or make connections between required knowledge or given information E - Apply concepts or procedures, or make connections, or make extended use of unfamiliar information **love** and **appreciation** not measured

Note:
a Categories named in Figure 4.6.

Table 4.3 Approach to the CBS used in this study (DNE, 1984a, 1984b) and their relationship to the major groups of the PET.

	Principles	Performance Expectations Taxonomy category [a]
The approach to the syllabus should embody the following principles:		
1.	Pupils should make their own observations of specimens and experiments.	B - Perform-Routine-Procedures
2.	Pupils should learn how to handle and set up apparatus correctly.	B - Perform-Routine-Procedures
3.	Organisms should be observed in their natural environment.	B - Perform-Routine-Procedures
4.	Constant emphasis should be placed upon facts being understood, interpreted and applied rather than being merely memori[z]ed.	A - Memorize C - Explain D - Analyze, or make connections between required knowledge or given information E - Apply concepts or procedures, or make connections, or make extended use of unfamiliar information

Note:
a Categories named in Figure 4.6.

The approach to the syllabus required by the CBS (Table 4.3) was that students be exposed to the routine procedures associated with science teaching (PET category B) and that students be required to understand, interpret and apply facts (PET categories C, D and E) rather that to simply memorize facts (PET category A). Therefore, the principles listed in the CBS as being of importance to how teachers approached the delivery of SC Biology to students, embraced all the PET categories.

4.5 Examples of Senior Certificate Biology examination questions classified using the Performance Expectations Taxonomy

The taxonomies or frameworks that have been described in this chapter were all developed out of a practical need to organize, or to make sense of, information about cognitive demand with respect to teaching and learning. Many of the taxonomies from the literature discussed in this chapter presented examples or vignettes which illustrated how they functioned, when they were described in the literature. Likewise, this section concludes by presenting specific examples of questions which were classified, according to implied cognitive demand, using the PET developed in this chapter.

In preparation for the validation of the PET, a set of questions were specifically prepared, that is, they did not appear in any of the questions papers analyzed in Chapters 5 and 6, to illustrate the practicality of the PET, and to illustrate the different categories of performance expectations recognized by the taxonomy to the raters involved its validation. These examples of questions used in preparation for the validation of the PET (Appendix 4.6), covered all categories of the PET, and drew on two specific content areas of the SC Biology syllabus, namely, the topics photosynthesis and cellular respiration.

To demonstrate how the post-validation model of PET was used in this study, the author compiled a comprehensive list of questions taken from the 111 question papers analyzed in Chapter 5 and Chapter 6, which describe each of the five categories of the PET. Included with each question is the rationale for its particular PET classification. The questions were selected to illustrate a range of different question types for each of the PET categories to give the reader authentic examples of how the PET was used in this study (Figures 4.14 to 4.18). In some cases the formatting of the questions/scorable events was changed to improve the readability and to allow comparisons of questions/scorable events, and only parts of some of the longer questions are shown. These questions were selected to indicate a range of the different types of questions in each of the PET categories *and* the range of different kinds of questions which appeared on SC Biology examinations, because many of the SC Biology question papers are no longer in the public domain.

4.6 Chapter summary

In Chapter 3 content standards were defined as consisting of two components, topic and cognitive demand. For this study, topic can be objectively identified from the CBS. This chapter describes the national policy requirements with respect to cognitive demand in South African SC Biology examinations during the period 1994 to 2007. Some measures of cognitive demand, including taxonomies of cognitive demand, which have been used in the literature are described. Reasons why none of these taxonomies are used in this study are discussed. The development, description and validation of a new taxonomy, the PET, for use in this study of SC Biology examinations, are described. The general objectives and principles of the CBS are mapped to the PET to show how the PET is infused across the CBS. Relationships between the PET, which was developed to analyze SC Biology examinations in this study, and the BTEO which had been used to set some of the SC Biology examinations are explored. Since the development of the PET arose from a practical need, the final section of this chapter provides some examples of questions taken from SC Biology examinations analyzed in this study. These example questions are classified according to each of the five PET categories Memorize, Perform-Routine-Procedures, Explain, Analyze and Apply. The PET was developed to identify the performance expectations that tasks, like examination questions, are intended to elicit, rather that specific cognitive processes for producing the performances. The PET has the potential to be used in other subjects and for analyses of summative assessment tasks. The PET forms a part of the methods which are used to explicate the content standards and the performance standards from the SC Biology examinations, the focus of this thesis. Development of the PET addresses part of research sub-question 1 (Chapter 1).

A. Orange Free State Education Department 1996 HG

1.2.10 Which ONE of the following waste products is secreted by the skin, lungs and kidneys?

A Carbon dioxide
B Mineral salts
C Urea
D Water (2)

6.5.1 Which ONE of the following combinations indicate the correct path whereby light enters the eye?

A Cornea Ÿ conjunctiva Ÿ pupil Ÿ lens Ÿ retina
B Conjunctiva Ÿ cornea Ÿ pupil Ÿ lens Ÿ retina
C Conjunctiva Ÿ cornea Ÿ lens Ÿ pupil Ÿ retina
D Conjunctiva Ÿ cornea Ÿ retina Ÿ pupil Ÿ lens (3)

Recall − therefore classified as Memorize.

B. Northern Cape Education Department 1997 HG

1.1.15 The volume of air in the alveoli increases when the é

A external intercostal muscles relax.
B diaphragm becomes more convex.
C ribs move downwards.
D external intercostal muscles contract. (3)

Recall − therefore classified as Memorize.

1.4 In the table below COLUMN A contains FOUR possible terms and column B is a statement. Choose TWO possibilities from COLUMN A which would be applicable [to] the statement in COLUMN B.

	COLUMN A	COLUMN B
1.4.1	A watery medium with a pH of 7 B hydrochloric acid C saliva D bile	The correct pH for enzyme activity in the cardiac portion of the stomach is created by é and é
Ź		
1.4.10	A NADP and water catabolized B ADP is formed C water is produced D NAD and oxygen catabolized	During respiration the following two processes take place[:] é and é

(2 marks each)

Recall − therefore classified as Memorize.

C. Gauteng Education Department 1998 HG

4.2.2 Name the illness that develops during an infection of the lower air passages which enter the lungs. (2)

Recall − therefore classified as category Memorize.

D. Independent Examinations Board 1998 HG

1.3.7 Name one common neurotransmitter substance that crosses the narrow gap [shown in diagram of a synapse which accompanies the question]. What happens to this substance once it has crossed the gap? (4)

Recall − therefore classified as Memorize.

Figure 4.14 continued on next page

E. Northern Province (now called Limpopo Province) Education Department 1998 HG

1.1 Each of the following questions comprises TWO items in COLUMN I and ONE statement in COLUMN II. Determine which item(s) refer to the statement and choose the code as follows:

A If the statement refers to (a) only
B If the statement refers to (b) only
C If the statement refers to BOTH (a) and (b)
D If the statement does NOT refer to either (a) or (b)

	COLUMN I	COLUMN II
1.1.1	(a) Pepsin (b) Trypsin	Hydrolyses proteins to form peptones
Ź		
1.1.12	(a) Oxyhaemoglobin (b) Dissolved in blood plasma	Represents the manner(s) in which oxygen is transported in the blood

(2 marks each)

For each question the student had to recall if each of the items (a) and (b) were linked to the corresponding statement or not. The complexity of these questions came from how the answer was to be represented not from the science of the answer − therefore classified as Memorize.

F. KwaZulu-Natal Education Department 2000 HG Paper 2

3.2.1 Compare the nervous system and the endocrine system by completing the following table. Write down the letters A to H and the missing information next to each letter.

Characteristic	Nervous system	Endocrine system
Nature of message	A	B
Message transported by	C	D
Area affected by the message	E	F
Speed of reaction	G	H

(8)

Although this question is making a comparison between two systems that was not required by the syllabus the student does not have to make the comparison − the characteristics for the comparison were given in the question. The student had to recall what these characteristics are for each of the systems independently. Therefore classified as Memorize.

G. National Department of Education 2004 HG Paper 1

1.1.4 The following substances are present in the colon

(i) Water
(ii) Some vitamins
(iii) Cellulose

Which ONE of the following shows the correct combination of substances absorbed by the walls of the colon?

A (i) only
B (i) and (ii) only
C (i) and (iii) only
D (i), (ii) and (iii) (2)

Recall of facts −therefore classified as Memorize.

Figure 4.14 Examples of questions/scorable events classified using the Memorize category of PET and explanations for their classification in this group.

A. Cape Education Department 1994 HG

6. é The average number of bubbles per minute emitted by the plant was counted and recorded at each distance. (See the table below.)

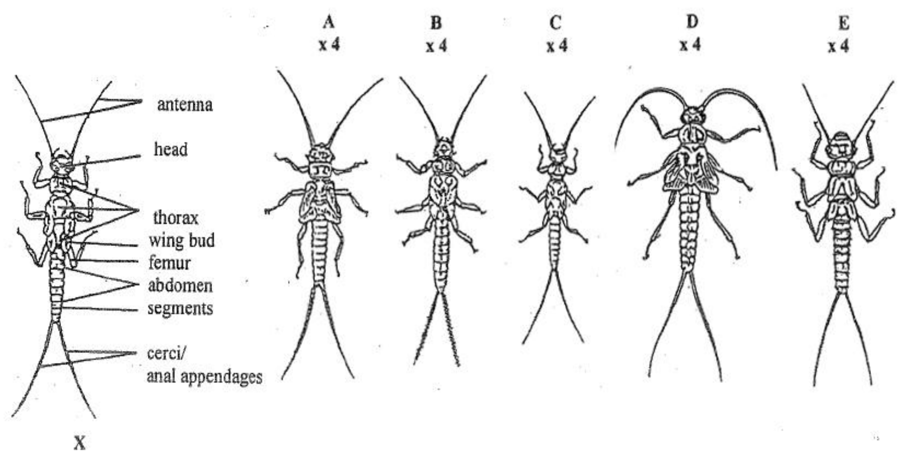
Number of bubbles set free per minute	99	95	55	30	15	10	5	2	1
Distance between lamp and plant in cm	5	10	15	20	25	30	35	40	45

6.1 Represent the results graphically on the graph paper which is supplied. Use the X-axis for ñDistance in cmò and the Y-axis for ñThe number of bubbles set free per minuteó (4)

Drawing a graph − therefore classified as Perform-Routine-Procedures

B. Natal Education Department 1995 HG

4.4 The following question tests your ability to observe accurately. Study the diagrams of these stonefly nymphs (young insects) which are found clinging to the underside of rocks in streams. The drawings show five different species of stonefly. Species X has been drawn a little larger but it is the same as one of the species A to E.

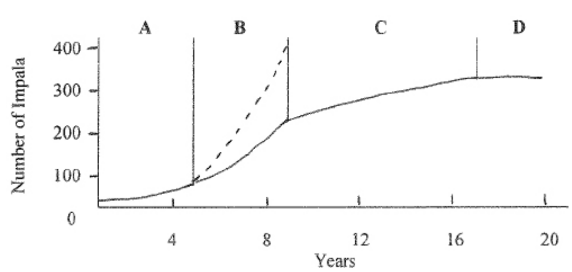


4.4.4 Give **three** visible differences (show in table form) between species B and species E. (6)

Making observations − therefore classified as Perform-Routine-Procedures

C. Northern Province (now called Limpopo Province) Education department 1997 HG

2.6 é Study the graph and answer the following questions.



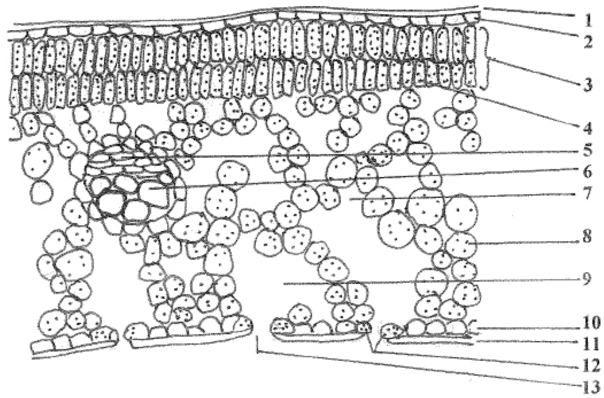
2.6.1 How many impala were there after twelve years? (1)

Reading a value from a graph − therefore classified as Perform-Routine-Procedures

Figure 4.15 continued on next page

D. National Examination Board 1995 HG

6. The following diagram shows a cross section (transverse section) through a dicotyledonous leaf as seen under a light microscope. The enlargement (scale) is shown below the drawing.



6.1 Measure the thickness of the leaf in the drawing and then, using the scale given, calculate the real thickness in millimetres. (Show your calculation.) (5)

Take a measurement and make a calculation − therefore classified as a Perform-Routine-Procedures.

E. Mpumalanga Education Department 1996 HG

2.2 Draw and label a diagram of the experimental and control apparatus and materials needed to show which gas is released during photosynthesis. Also indicate the results of the experiment. (10)

Draw a learned experiment − therefore classified as a

F. Gauteng Education Department 1998 HG

4.5.2 Draw and label a simplified sketch of a coiled sweat gland as found in mammalian skin showing the structures which are involved in thermoregulation. (4)

The candidate is required to draw a learned structure − therefore classified as a 'Perform-Routine-Procedures.

G. Western Cape Education Department 1997 HG

6.1 é for every 10g of [lactic] acid the body must absorb 1.7 litres of oxygen in order to break the acid down é

6.1.6 From the data in the passage determine the amount of oxygen that the body of an athlete must absorb if 120g of lactic acid has accumulated in his muscle tissue. Show your calculations. (2)

Do a calculation − therefore classified as Perform-Routine-Procedures.

H. National Education Department 2004 HG Paper 2

2.2 é Study the table below of the results obtained, and answer the questions that follow.

Time (minutes)	1	2	3	4	5
Distance travelled by air bubble along capillary tube (mm)	4	11	18	30	22

2.2.3 Draw a bar graph to present the data shown in the table. (11)

Draw a graph as instructed − therefore classified as Perform-Routine-Procedures.

Figure 4.15 Examples of questions/scorable events classified using the Perform-Routine-Procedures category of PET and explanations for their classification in this group.

A. Cape Education Department 1995 HG

1.1.4 Saliva was mixed with a cooked starch suspension and divided into four equal parts. Each part was kept at a certain temperature for 15 minutes. The amount of maltose formed at each temperature was measured and tabulated below.

Temperature (^oC)	20	40	60	80
Amount of maltose formed (mg)	60	90	30	2

At what temperature would you expect to find the most starch after 15 minutes?

A 20 ^oC
B 40 ^oC
C 60 ^oC
D 80 ^oC (2)

Show an understanding of the relationship between saliva, starch and maltose, all from one topic (biological compounds or human nutrition) − therefore classified as Explain.

B. House of Representatives Education Department 1995 HG

2.1 The following table shows the amount of energy (in kJ) in three different cooldrinks.

kJ per litre	Cooldrink
100	diet cola
80	orange soda
540	cola

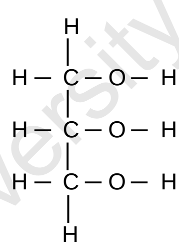
2.1.3 Most of the energy you need for dancing you need for dancing is obtained from sucrose (disaccharide) in the cooldrink. Before you can use the energy you have to digest the cooldrink. Explain how sucrose is digested in the small intestine.

The first two sentences only give a background to this question and were not relevant to the question. Explain a learned process − how sucrose is digested. Therefore classified as Explain.

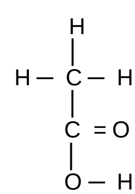
C. Eastern Cape Education Department 1997 HG

2.1 Study the structural formulae given below and answer the questions that follow.

(i)



(vi)



2.1.2 Use structures (i) and (vi) to show how an ester bond is formed. (4)

Explain a learned concept − how ester bonds are formed. Therefore classified as Explain.

D. Eastern Cape Education Department 1998 HG

4.2 ñAdrenalin prepares the body to cope with stress situationsò Explain this statement. (6)

Explain a learned function − what adrenalin does that enables the body to cope with stress.. Therefore classified as Explain.

E. Northern Cape Education Department 1999 HG

3.1.4 Explain fully **how** and **why** an increase in the breathing rate is brought about in humans. (4x2) = (8)

Explain a learned facts and concepts − why humans need to increase breathing rate and how humans increase breathing rate. Therefore classified as Explain.

Figure 4.16 continued on next page

F. Independent Examinations Board 1997 HG

1.3.5 Briefly explain why there is a gap between the two neurons. (2)

Explain a learned concept − the relationship between the structure of a synapse and its function. Therefore classified as Explain.

G. KwaZulu-Natal Education Department 1999 HG

6.2 The table below shows the food value of a school lunch eaten by a 16 year old girl.

Food eaten	Protein[s] (g)	Carbohydrates (g)	Fats (g)	Iron (mg)	Vitamin C (mg)
Sausages	9	5	24	1	0
Chips	8	70	20	2	20
Baked beans	10	20	1	3	4
Apple pie	5	60	25	1	1
Ice cream	2	20	12	0	0
Fizzy drink	0	30	1	0	0

6.2.4 A lot of energy comes from carbohydrates and fats. Name the food in this meal which would give her the LEAST amount of energy. (2)

Show evidence of basic understanding − to find the answer the candidate needed to look up in the table the food with smallest combination of carbohydrates and fat (given in the question) - therefore classified as Explain.

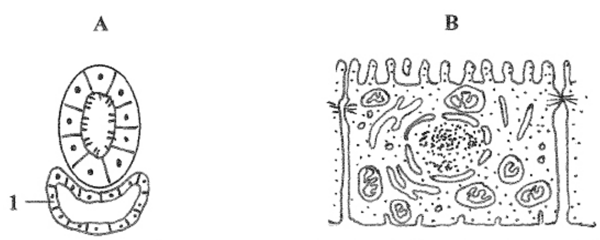
H. National Education Department 2007 SG Paper 2

3.2.4 Explain why a temperature of 25[.]0 ^{o}C will lead to death [in humans]. (3)

Explain a learned concept − the effect of suboptimal temperatures on the human body. Therefore classified as Explain.

I. National Education Department 2007 HG Paper 2

1.5 The diagram below represents a transverse section through the proximal convoluted tubule (A) and a single epithelial cell (B) of the same tubule in a human nephron. Study the diagram and answer the questions that follow.



Transverse section of proximal convoluted tubule and a single epithelial cell

1.5.1 Explain TWO visible adaptations in cell B that enable it to perform its function efficiently. (4)

Observe ("visible") and explain an adaptation that has been learned – cannot be simply recalled because question is directly related to the diagram given in the question. Therefore classified as Explain.

Figure 4.16 Examples of questions/scorable events classified using the Explain category of PET and explanations for their classification in this group.

A. Transvaal Education Department 1994 HG

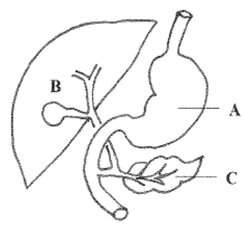
1.1.4 Name the macro-nutrient which plays a role in the transfer of energy in the organelle labelled Y.

A Iodine
B Magnesium
C Phosphorus
D Potassium (2)

Students learned in the biochemistry section of the Grade 12 (Standard 10) syllabus that phosphorus is necessary for the formation of ADP and ATP. In the section on cellular respiration, also Grade 12 the role of ADP and ATP in energy capture and release was required. Mitochondria as cellular organelles was part of the Grade 10 (Standard 8) syllabus which concerned the structures of eukaryotic cells which was not required to be revised in Grade 12. To answer this question the student had to make connections between different learned knowledge (familiar context) − therefore classified as Analyze.

B. Orange Free State Education Department 1995 HG

3.1 Study the diagram and answer the questions that follow.



3.1.3 Explain why C is known as both an endocrine and an exocrine gland. (2)

Connect the structure and function of the pancreas learned in two Grade 12 (Standard 10) chapters, human nutrition and co-ordination. Therefore classified as Analyze.

C. House of Representatives Education Department 1995 HG

2.1 The following table shows the amount of energy (in kJ) in three different cooldrinks.

kJ per litre	Cooldrink
100	diet cola
80	orange soda
540	cola

2.1.1 Which cooldrink would be the best to ensure that you stay slim? (1)

Deductive reasoning: link the relationship between kJ and size and link this insight to the data given in the graph. Therefore classified as Analyze.

D. Eastern Cape Education Department 1998 HG

2.5.3.1 Insulin cannot be taken by mouth because it would be digested in the digestive system. Give the names of TWO enzymes that would digest insulin. (2)

Link that insulin, because it is a hormone, is a protein and that digestive enzymes digest proteins. Therefore classified as Analyze.

E. Northern Province Education Department 1999 HG

3.4.1 Give ONE difference between this [nervous] system and the endocrine system. (2)

The student has learned about the nervous and endocrine systems (i.e., familiar context) but the syllabus does not require the student to make comparisons of the two systems. Therefore classified as Analyze,

Figure 4.17 continued on next page

F. Independent Examinations Board 1998 HG

3.1 Although their functions are completely different, mitochondria and chloroplasts have some similarities in structure. List any THREE structural similarities between mitochondria and chloroplasts. In each case, give a reason concerned with their function to explain the similarity. (9)

Perform a comparison between mitochondria and chloroplasts that is not required by the syllabus. Therefore classified as Analyze.

G. KwaZulu-Natal Education Department 2000 HG Paper 1

2.2.4 Tabulate THREE differences between the absorption of lipids and glucose. (6)

These differences were not required by the syllabus − the students needed to make the connection. Therefore classified as Analyze.

Had marks been awarded to the construction of a table a separate scorable event would have recorded these marks as category 'Perform-Routine-Procedures.

H. National Department of Education 2003 SG Paper 1

3.2 The table below shows the details of a meal eaten by a Grade 10 learner.

Food eaten	Carbohydrate (g)	Protein (g)	Fat (g)	Vitamin C (mg)	Iron (mg)
Sausages	5	9	24	0	1
Fried potato chips	70	8	20	20	2
Beans	20	10	1	4	3
Fruit pie	60	2	25	1	1
Ice cream	20	0	12	0	0

3.2.5 (iii) Name TWO foods from the table that provide the **most** energy for the learner. (2)

Recall that large amounts of energy come from carbohydrates and fats, before the food with the biggest combination of carbohydrates and fats could be located in the table. To successfully answer this question the candidate needed to made a connection between what he/she learned and what was given in the question. Therefore classified as Analyze.

I. National Department of Education 2004 HG Paper 2

5.1 Study the following diagram[s] and answer the questions that follow.

é

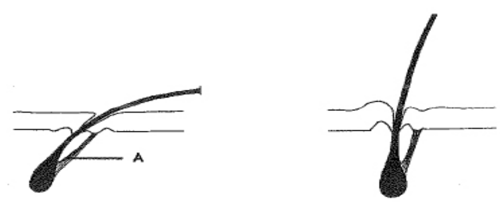


Diagram 2 Diagram 3

5.1.5 Diagram 2 shows the skin of person X and Diagram 3 shows the skin of person Y. Assuming all other factors are constant, explain which person (X or Y), will probably produce more dilute urine? (5)

Recall that when it is cold people produce more dilute urine and then from work learned in thermoregulation section decide which person is colder. Linking two different sections of work. Therefore classified as Analyze.

Figure 4.17 Examples of questions/scorable events classified using the Analyze category of PET and explanations for their classification in this group.

A. House of Delegates Education Department 1994 HG

6.1 [Long passage which discussed grouse population structure and population dynamics in Scotland. The passage names no density-independent factor which might have affected the size of the grouse population]

6.1.2 Name ONE density-independent factor which may affect the size of the grouse population. (2)

Recall what a density-independent factor was and then recognise it in descriptions of the niche of grouse described in the passage (unfamiliar context). Therefore classified as Apply.

B. Gauteng Education Department 1996 HG

1.1.11 Which of the following statements regarding the washing powder [in the accompanying diagram] is CORRECT?

The washing powder é

A will remove blood and meat stains.
B will function effectively in boiling water.
C functions best at a low pH.
D will increase the activation energy of the process.

(2)

WIWI-ROBO WASHING POWDER
ENZYME ACTIVE
Gets rid of the most stubborn <u>protein stains</u>!

SAVE ELECTRICITY!
Wash in <u>cold water</u>
(Max. 20^{o}C)

TAKE CARE!
Do not use in <u>acidic</u> water

Knowledge of enzymes (syllabus) was necessary to understand what was written about a particular washing powder (new context) – therefore classified as Apply.

C. Cape Education Department 1995 HG

4.3.9 Why is it possible to treat a diabetic successfully by administering insulin in the form of a nasal spray? (2)

Knowledge of insulin used in an unfamiliar context. Therefore classified as Apply.

D. Independent Examinations Board 1998 HG

1.3.8 Strychnine is a chemical substance that stops the breakdown of the neurotransmitter when it reaches the postsynaptic membrane (D on diagram). What effect would strychnine have on the body if it reached the synapses between the [c]entral [n]ervous [s]ystem and the motor neurons linked to muscles? (2)

Syllabus required that students learn only that impulses move across synapses chemically (as opposed to electrically within a neuron). They need to infer from this information what would happen if information from the same impulse moves backwards and forwards across the same synapse (unfamiliar context). Therefore classified as Apply.

E. Orange Free State Education Department 2000 HG

5.6.2 Why are hedges with green leaves more often pruned than hedges with variegated leaves?

A The net primary production of green leaves is higher than that or variegated leaves.
B Plants with green leaves are stronger.
C Plants with green leaves have a higher metabolic rate.
D A higher flow of energy occurs in a plant with green leaves. (3)

Required to put together learned information about productivity, photosynthesis, cellular respiration and plant growth in an unfamiliar context. Students would have learned about variegated leaves in the section on photosynthesis. Therefore classified as Apply.

Figure 4.18 continued on next page

F. Independent Examinations Board 1999 HG

5.4 Egg albumen is a water soluble protein that makes up much of the protein store in an egg. Explain why it turns solid and white when you heat an egg but does not become clear and runny again when you cool it. (5)

Required to make applications outside of what is required by the syllabus. Therefore classified as Apply.

G. National Department of Education 2001 SG Paper 1

3.3.4 Give ONE reason for each of the following:

(a) Breathing through the nose rather than through the mouth[.] (2)
(b) Talking stops when one is about to swallow some food[.] (2)

Required to make applications outside of what is required by the syllabus. Therefore classified as Apply.

H. National Department of Education 2005 SG Paper 2

3.2 [A short text described a ñvery drunkò Martin who was hospitalized. In the hospital he was given a lot of liquid after which he started to feel better]

Note:

- When the muscles are badly bruised, protein in the muscles breaks down. The products enter the kidney tubules and can block them.
- When a person drinks alcohol, it is absorbed into the blood, then transported to the hypothalamus, where it inhibits the release of ADH.

3.2.1 Explain why Martin was dehydrated? (4)
3.2.2 Account for the presence of a few drops of dark urine after being given a lot liquid. (3)
3.2.3 Explain why Martinõs body started to swell the next day. (3)

Synthesize ideas from a number of different, unfamiliar sources what was learned. Therefore classified as Apply.

I. National Department of Education 2006 HG Paper 1

3.2 [A description of an experiment using two groups of puppies fed on different diets. A table contrasting the diets was given as well as table which showed the average mass of each group over a period of 42 days]

3.2.2 Why were puppies of the same litter used in the experiment? (2)

Required an understanding of an appropriate experimental design in an unfamiliar context. Therefore classified as Apply.

J. National Department of Education 2006 HG Paper 1

5.3 [Short text on avian/bird flu (not part of the syllabus)]

5.3.3 Indicate whether bird flu should be considered a density-dependent or density-independent factor. Give a reason for your answer. (3)

Using concepts learned in population dynamics to understand the spread of avian flu (unfamiliar context). Therefore classified as Apply.

Figure 4.18 Examples of questions/scorable events classified using the Apply category of PET and explanations for their classification in this group.

CHAPTER 5

METHODS FOR THE ANALYSIS OF QUESTION PAPERS AND CANDIDATES' ANSWER SCRIPTS

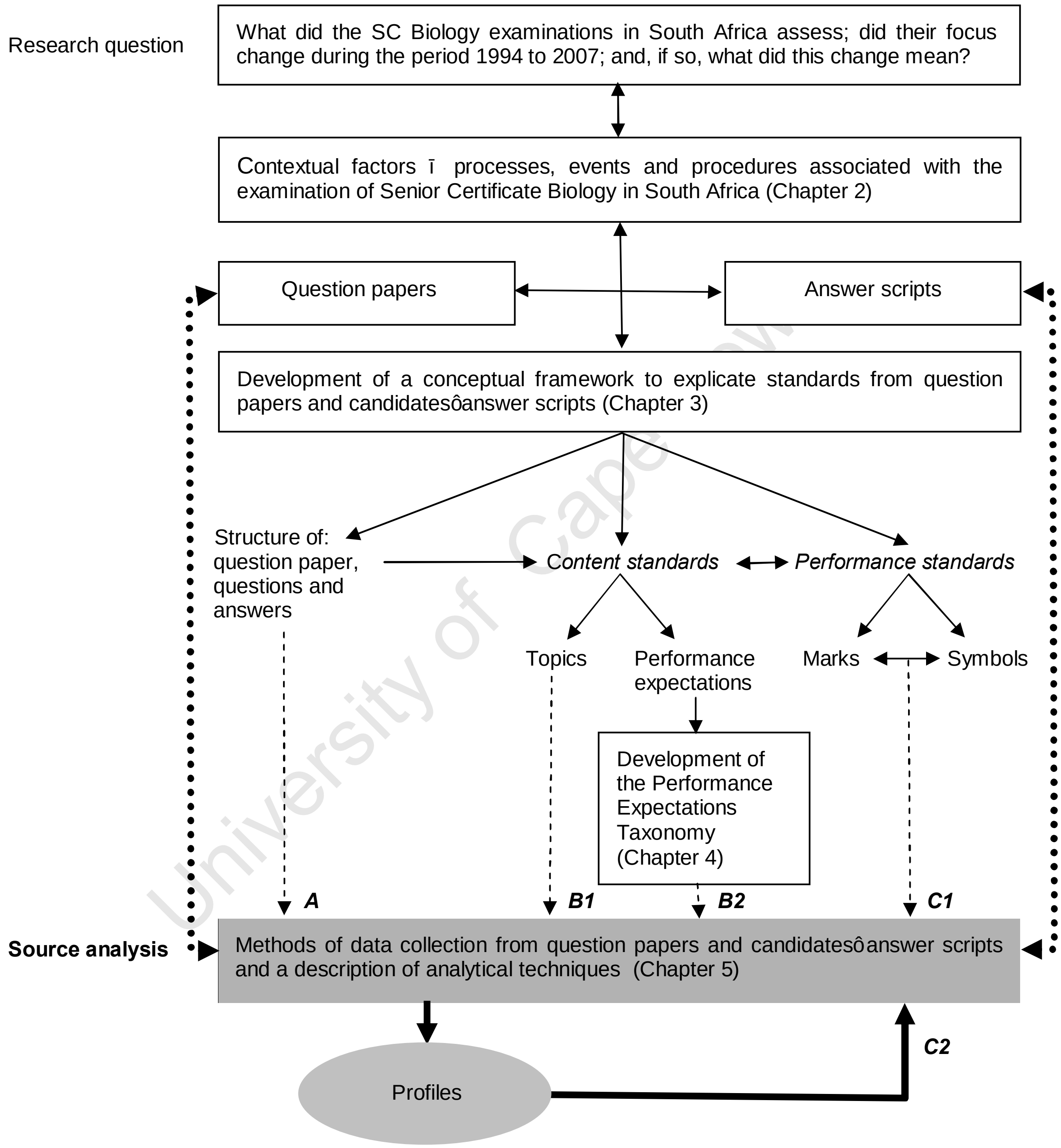


Figure 5.1 The relationship between Chapter 5 and the previous chapters.

Assessments like the SC Biology examinations, “are powerful instruments for educational change” (Shavelson & Huang, 2003, p. 11), provided that their results are interpreted in a meaningful way and each assessment, represents an interrelationship between its intended purpose and the context in which it is used (Mislevy, 1995). In Chapter 2, the purpose of the SC Biology examination, the context in which they functioned, and the policies which guided the examinations were described in detail. South Africa does not offer an explicitly standards-based curriculum, and so what competencies the SC examinations measure or don’t measure, in terms of standards, is unclear. Without information about competencies, society cannot interpret students’ achievement in examinations (Nitko & Brookhart, 2007), nor can society begin to understand what changes in student achievement between years might mean because a “cost of not having national standards is the cost of non-comparability” (Noah, 1989, p. 18).

Therefore, it was necessary to develop a conceptual framework to connect the research question (i.e., What did the SC Biology examinations in South Africa assess; did their focus change during the period 1994 to 2007; and, if so, what did this change mean?), with the sources of data, SC Biology examination question papers and candidates’ answer scripts, available for study, using standards as a lens. In Chapter 3, when the conceptual framework was developed it was argued that in the absence of explicit standards, as is the case in South Africa, different kinds of validity evidence can be used to explicate the standards inherent in the SC Biology examinations. Explicating these standards then allows comparisons to be made of different SC Biology examinations.

Validity confers meaning to student performances in tests and examinations (Cizek et al., 2008). The conceptual framework developed to guide this study identified three strands of validity evidence that are important in understanding, and comparing, student performance in the SC Biology examinations. These three strands are the structural aspects of question papers, called the *structural strand* (Figure 5.1[*A*]); the *content strand* (generates content standards) (Figure 5.1 [*B1* and *B2*]); and the *performance strand* (generates performance standards) (Figure 5.1 [*C1* and *C2*]). It is the interplay between these three strands that gives an explicit meaning to a candidate’s marks in a specific examination (question paper and answer scripts). Therefore, these three strands collectively describe the standards of an examination, and allow comparisons to be made between SC Biology examinations within and between years.

Each strand of different components which need to be determined for each examination. Of these components the structure of a questions paper, the questions which are and the expected answers to these questions (Figure 5.1 [*A*]*)*, and the topics covered by a question (Figure 5.1 [*B1*], can all be subjectively measured. The remaining component, of the content strand recognized as an important source of evidence about content standards in this study, is cognitive demand (Figure 5.1 [*B2*]). A review of the literature indicated that determining the cognitive demand of assessments has not been done consistently nor reliably (Chapter 4). Chapter 4 described the development of an instrument, the PET, to determine the cognitive demand of examination questions in this study. Given the various different definitions in the literature of performance standards, an argument was made in Chapter 3 to use the following definition for this study: performance standards are descriptions of how students demonstrate their proficiency or mastery in content standards. As marks achieved in a question or sub-question reflect a student's level of performance (i.e., their competence), marks are an integral part of the working definition of performance standards (Figure 5.1 [*C1*]). As performance standards are related to specific content standards, the content standards have to be identified before the performance standards can be generated and interrogated. In this study the identification of the content standards forms part of the profile of a question paper described below (Section 5.2.1.3). The profile generated for each question paper (Figure 5.1 [*C2*]) is then used to explicate the performance standards for that specific question paper.

This thesis is essentially a study involving the analysis of documents and involves multiple methods some of which were discussed in previous chapters. The first part of this chapter concerns the sources of data, that is, examination question papers and candidates' answer scripts, from which standards of the SC examinations can be elicited. Descriptions of both data sources include the limitations and delimitations associated therewith. This chapter concludes with a description of analytical tools used to generate, and to compare, the content standards and the performance standards implicit in both the data sources. The generated standards are described, analyzed and compared in Chapter 6, and provide the evidence used to answer the research question and research sub-questions, 2 to 7, in Chapter 7. This chapter completes the operationalization of the conceptual framework developed in Chapter 3 and attention to research sub-question 1 (i.e., What are *educational standards,* and how might they be used to describe and compare SC Biology examination question papers and candidates' answer scripts?).

5.1 Data sources

A detailed description of the policies, processes and procedures associated with SC Biology examinations is given in Chapter 2 (Section 2.1). Where necessary or appropriate, some of this information is repeated in this chapter to help the reader to understand the nuances/idiosyncrasies of the South African education practices which influence standards.

The remainder of this thesis involves the analyses and interpretations of data extracted from two sources of data, the SC Biology examination question papers, examinees' answer scripts for some of these examinations obtained in the examinations. Data derived from the examination question papers will be used to generate the content standards of each SC Biology examination and data from the individuals' actual responses to the test items of the specific examinations will be used to generate the performance standards of four of these examinations (Chapter 6).

5.1.1 Senior Certificate Biology examination question papers

The end-of-year South African SC Biology examinations administered in 1994 to 2007 were used in this study.[134] English versions of the examination question papers from all the government-controlled examining bodies and the IEB (Chapter 2, Table 2.4) were included.[135] Higher Grade examination question papers for the entire 14 years (i.e., 1994 to 2007) were analyzed and SG examination question papers were analyzed, for the seven-year period 2001 to 2007.[136] The sources of examination question papers used in this study were Scheltema and Myburgh, (1995, 1996, 1997, 1998, 1999, 2000, 2001),[137] the national DoE (DoE, 2001e, 2001f, 2001g, 2001h, 2002e, 2002f, 2002g, 2002h, 2003b, 2003c, 2003d, 2003e, 2003f, 2004e, 2004f, 2004g, 2004h, 2004i,

134 Supplementary examinations written by candidates who failed the examination in the previous year, or who were unable to write examination in the previous year, were not included in this study.

135 Question papers used by the BCVO were omitted from this study as they were set only in Afrikaans. Question papers from the SoT were unobtainable.

136 A complete set of 1994- 2007 SC SG Biology examination question papers for the period 1994 to 2000 was unobtainable.

137 Each year from 1994 to 2004 an annual one-day meeting was held to discuss the Biology examinations of the previous year. This meeting of representatives of all the examining bodies was convened by the certification body of the time, that is, SAFCERT or Umalusi, and chaired by the external moderator(s) appointed by them. As one of these external moderators during this time I collected copies of each of the SC Biology examination question papers discussed at these meetings. Each year I checked that each examination paper that appeared in a compilation of questions papers (the Bios HG series) for that year was the same as that of the original question papers and retained the smaller compilation for reference purposes. The Bios HG books were for sold in bookstores. There was no Bios SG series.

2004j, 2005e, 2005f, 2005g, 2005h, 2006c, 2006d, 2006e, 2006f, 2007a, 2007b, 2007c, 2007d) and the IEB (IEB, 2001a, 2001b, 2002a, 2002b, 2003a, 2003b, 2004a, 2004b, 2005a, 2005b, 2006a, 2006b, 2007b, 2007c).[138]

In the period 1994 to 2000, both HG and SG question papers consisted of one three-hour question paper for all examining bodies. From 2001 until 2007, the national DoE HG and SG question papers were divided into two papers of two hours each per grade, but the number of IEB HG and SG question papers remained unchanged. In some analyses, data from Paper 1 and Paper 2 are combined for comparative purposes.

The 111 individual question papers which were analyzed in this study and their sources are listed in Appendix 5.1.

5.1.2 Candidates' answer scripts

In approximately the middle of the year following an examination, when Umalusi has issued the Senior Certificates, candidates' answer scripts are destroyed by examining bodies.[139] This practice has meant that samples of candidates' answers could not be analyzed for all the examination question papers used in this study. For the years 2005 and 2006, stratified random samples were selected from the education department of one province, the WCED (Section 2.2.4), for analysis of candidate performance.[140] In 2005, a request was made for 100 [141] candidates' answer scripts, randomly selected, from each of the ten symbols (i.e., A to H) which candidates could obtain, for each of the four papers, that is, HG Paper 1, HG Paper 2, SG Paper 1 and SG Paper 2. In 2006 the sampling procedure was changed so that the Paper 1 and Paper 2 from each of the same 100 randomly selected candidates, per overall outcome symbol, was obtained for both HG and SG.

---

138 For the period 2001 to 2007, original question papers from the national DoE and the IEB were used. Although I obtained these question papers in my capacity as a SAFCERT/Umalusi external moderator, these question papers are all in the public domain.

139 From 2010 the IEB has retained a sample of the answer scripts of candidates who write their examinations.

140 Some of the selected scripts were from students who wrote the examination in Afrikaans. The English and Afrikaans versions of the 2005 and 2006 question papers were deemed to be testing the same construct by the examination panel who set the paper, by Umalusi external moderators and by public comment. Those questions considered to be different are noted in Chapter 6.

141 For some symbols with fewer than 100 candidates, all candidates with that symbol were selected. Also, and in some instances marks were incorrectly totalled before selection.

Sampling the candidates answer scripts in a sub-optimal way in 2005 and appropriately in 2006, limits the extent of statistical comparisons which can be made, or the evidence for generalizations which might be made from the analyses of the data obtained for the two years. However, having 2006 Paper 1 and Paper 2 from the same candidates made the 2006 data richer, because comparisons could be made of the performances of the same candidate in both papers.

No distinction was made between English and Afrikaans answer scripts in the sample because the English and Afrikaans memoranda used to mark the scripts were standardized prior to marking. The resulting memoranda were considered by bilingual Biology experts to be completely equivalent for marking purposes.[142] Unfortunately, the author did not have access to the personal data of the students whose scripts were used in this study. This limitation meant that students who wrote the 2005 and 2005 Biology SC examinations in a language other than their home language could not be identified.

The distributions of the sampled 7 553 candidates answer scripts across the symbols together with the symbol distributions for all candidates who wrote the examinations are given in Table 5.1 (2005) and in Table 5.2 (2006). For four of the eight examinations concerned, there were invalid questions (Table 5.3) for which every candidate was awarded notionally full marks in the examination.[143] The symbols for these four examinations were thus recalculated with the marks awarded for invalid questions[144] removed (Table 5.1 and Table 5.2). Recalculated symbols decreased the numbers of candidates who passed the Grade on which they wrote and decreased the number of candidates who attained higher symbols.[145] The DoE practice of awarding full marks to candidates for invalid questions therefore inflates the marks of students. The scripts were retained in the original categories to which they were assigned within the WCED, except where addition

142 The memorandum standardization process takes into account any subtle differences which might exist between the English and Afrikaans translations of the question papers.

143 All students were awarded the full three marks for one question (2006 SG Paper 1 Question1.4.2) even if their answers were incorrect. This question was not considered to be invalid at the memorandum meeting, nor could any reason for this practice be found. Therefore, this question was not considered to be invalid in the student analyses reported in the Chapter 6.

144 Any question deemed to be flawed during the memorandum discussion meeting was considered to be invalid.

145 The practice of awarding full marks for answers given to invalid questions during the period of this study continues currently (2011). What this study shows is that invalid questions should be removed from the aggregate if the aggregate mark is to be accepted as representative of what a candidate demonstrates they know (Chapter 6).

Table 5.1 Total counts of WCED candidates and selected scripts analyzed for the 2005 SC Biology examinations. Shaded cells indicate candidates who failed the Grade they wrote.

A. HG Paper 1

Symbol	A	B	C	D	E	F	FF	G	GG	H	Total	Pass HG	Fail HG	Pass SG [a]	Pass SG [b]	Fail HG/ SG [c]
Total number of candidates	521	1 194	1 567	1 639	1 526	788	333	361	274	249	8 452	6 447	2 005	1 121	1 482	523
Number of scripts in sample	100	101	100	101	98	99	100	102	98	102	1 001	500	501	199	301	200
Number of scripts which move to a lower symbol in adjusted sample [d]	8	7	7	11	9	15	22	20	16	0	115					
Number of scripts in adjusted sample	92	102	100	97	100	93	93	104	102	118	1 001	491	510	186	290	120

Note:

a HG failures: symbols F and FF converted to SG passes for SC with ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007.
b HG failures: symbols F, FF and G converted to SG passes for SC without ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007.
c Did not pass on either HG or SG.
d Three marks credited to all candidates for invalid question removed from analyses of sample.

Table 5.1 continued

B. HG Paper 2

Symbol	A	B	C	D	E	F	FF	G	GG	H	Total	Pass HG	Fail HG	Pass SG [a]	Pass SG [b]	Fail HG/ SG [c]
Total number of candidates	327	961	1 534	1 778	1 840	863	370	381	237	161	8 452	6 440	2 012	1 233	1 614	398
Number of scripts in sample	102	99	100	100	100	102	98	101	102	100	1 004	501	503	200	301	202
Number of scripts which move to a lower symbol in adjusted sample [d]	8	11	12	14	13	25	58	43	38	0	222					
Number of scripts in adjusted sample	94	96	99	98	101	90	65	116	10	138	1 004	488	516	155	271	245

Note:

a HG failures: symbols F and FF converted to SG passes for SC with ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007.

b HG failures: symbols F, FF and G converted to SG passes for SC without ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007.

c Did not pass on either HG or SG.

d Five marks were credited to all candidates for invalid question removed from sample data.

C. HG Paper 1 and Paper 2 combined

Symbol	A	B	C	D	E	F	FF	G	GG	H	Total	Pass HG	Fail HG	Pass SG [a]	Pass SG [b]	Fail HG/ SG [c]
Total number of candidates	366	1 063	1 552	1 786	1 697	825	379	353	253	178	8 452	6 464	1 988	1 204	1 557	431

Note:

a HG failures: symbols F and FF converted to SG passes for SC with ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007.

b HG failures: symbols F, FF and G converted to SG passes for SC without ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007.

c Did not pass on either HG or SG.

Table 5.1 continued

D. SG Paper 1

Symbol	A	B	C	D	E	F	FF	G	GG	H	Total	Pass SG	Fail SG	Pass LG [a]	Fail SG/ LG [b]
Total number of candidates	111	444	1 044	2 007	3 145	2 436	1 853	2 188	2 321	2 630	18 179	9 187	8 992	4 041	4 951
Number of scripts in sample	81	101	99	100	101	100	98	103	100	99	982	582	400	201	199

Note:
a SG failures: symbols FF and G converted to LG passes for SC without ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007.
b Did not pass on either SG or LG.

E. SG Paper 2

Symbol	A	B	C	D	E	F	FF	G	GG	H	Total	Pass SG	Fail SG	Pass LG [a]	Fail SG/ LG [b]
Total number of candidates	6	48	221	726	1 733	2 026	1 810	2 549	3 242	5 818	18 179	4 760	13 419	4 359	9 060
Number of scripts in sample	6	47	100	100	100	101	100	99	99	102	854	454	400	199	201

Note:
a SG failures: symbols FF and G converted to LG passes for SC without ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007.
b Did not pass on either SG or LG.

F. SG Paper 1 and Paper 2 combined

Symbol	A	B	C	D	E	F	FF	G	GG	H	Total	Pass SG	Fail SG	Pass LG [a]	Fail SG/ LG [b]
Total number of candidates	25	131	499	1 263	2 636	2 597	1 746	2 719	2 751	3 812	18 179	7 151	11 028	4 465	6 563

Note:
a SG failures: symbols FF and G converted to LG passes for SC without ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007.
b Did not pass on either SG or LG.

Table 5.2 Total counts of WCED candidates and selected scripts analyzed for the 2006 SC Biology examinations**.** Shaded cells indicate candidates who failed the Grade they wrote.

A. HG Paper 1 and Paper 2 combined

Symbol	A	B	C	D	E	F	FF	G	GG	H	Total	Pass HG	Fail HG	Pass SG [a]	Pass SG [b]	Fail HG/ SG [c]
Total number of candidates – Paper 1 & Paper 2	483	1 087	1 578	1 786	1 764	931	362	368	195	129	8 683	6 698	1 985	1 293	1 661	324
Number of scripts in sample – Paper 1 & Paper 2	94	92	99	98	97	98	99	100	100	98	975	480	495	197	297	198
Total number of candidates - Paper 1	1 183	1 517	1 727	1 763	1 377	598	189	180	71	78	8 683	7 567	1 116	787	967	149
Number of scripts in sample – Paper 1 [d]	151	110	94	93	149	127	59	85	54	53	975	597	378	186	71	107
Total number of candidates - Paper 2	200	713	1 193	1 597	1 777	1 130	545	645	478	405	8 683	5 480	3 203	1 675	2 320	883
Number of scripts in sample – Paper 2	46	89	83	92	98	76	56	108	126	201	975	408	567	132	240	327
Number of scripts which move to a lower symbol - Paper 2 adjusted sample [e]	1	5	14	4	9	15	18	41	20	0	133					
Number of scripts – Paper 2 adjusted sample	45	85	74	102	93	70	53	85	147	221	975	399	576	123	208	368

Note:

a HG failures: symbols F and FF converted to SG passes for SC with ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007.

b HG failures: symbols F, FF and G converted to SG passes for SC without ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007

c Did not pass on either HG or SG.

d No invalid questions, therefore no adjustment to sample size.

e Four marks credited to all candidates for invalid questions removed from analyses of sample.

Table 5.2 continued

B. SG Paper 1 and Paper 2 combined

Symbol	A	B	C	D	E	F	FF	G	GG	H	Total	Pass SG	Fail SG	Pass LG [a]	Fail SG/ LG [b]
Total number of candidates – Paper 1 and Paper 2	10	98	435	1 266	3 130	3 076	2 209	3 364	2 838	1 984	18 410	8 015	10 395	5 573	4 822
Number of scripts in sample aggregate sample	9	96	98	98	98	95	96	95	98	98	881	494	387	191	196
Total number of candidates – Paper 1	57	380	1 139	2 545	4 177	3 304	2 255	2 283	1 602	668	18 410	11 602	6 808	4 538	2 270
Number of scripts in sample – Paper 1	43	126	107	99	138	122	63	84	67	32	881	635	246	147	99
Total number of candidates – Paper 2	7	45	171	583	1 643	2 045	1 942	2 799	3 750	5 425	18 410	4 494	13 916	4 741	9 175
Number of scripts in sample – Paper 2	6	41	89	82	98	92	42	90	130	211	881	408	473	132	341
Number of scripts which move to a lower symbol – Paper 2 adjusted sample [c]	0	5	6	5	8	11	11	10	26	0	82				
Number of scripts –Paper 2 adjusted sample	6	36	88	83	95	89	42	91	114	237	881	397	484	133	351

Note:

a SG failures: symbols FF and G converted to LG passes for SC without ME (DoE, 1997b). This process was to have been curtailed at the end of 1999 but continued until 2007.

b Did not pass on either the SG or on the LG.

c Two marks credited to all candidates for invalid questions removed from analyses of sample.

Table 5.3 Invalid questions removed from script analyses.

Year	Question paper	Question no.	No. marks	Reason question invalid
2005	Paper 1 HG	2.2.4	3	Shading did not print in all copies of the graph used in this question
	Paper 2 HG	3.2.8	5	Question is scientifically incorrect
2006	Paper 2 HG	1.3.5 5.1.7	2 2	Requires recall of facts outside of syllabus Question is scientifically incorrect
	Paper 2 SG	1.4.7	2	Question is scientifically incorrect

errors had been made by the WCED. Invalid questions were removed for analyses of the candidate performances for the four papers which contained them, except where noted in the Chapter 6.[146]

5.1.3 The representivity and reliability of the sample of candidates' answer scripts

In order to determine how generalizable the findings of this study based on analysis of the sampled scripts may be, it was necessary to determine how representative the sampled findings are by comparison with corresponding data for the total population of candidates who wrote each examination. The representivity of the samples of scripts and the populations for which they correspond was established by comparison of the sample data with WCED data[147] for the entire cohort of candidates who wrote the eight examinations (Table 5.4). A comparison of the mean marks achieved by the candidates whose scripts comprised the eight samples of scripts and the mean marks achieved by all the candidates of each of the cohorts represented, must be made in the context of the distributions of the total marks achieved by each cohort of candidates (Tables 5.1

146 For example, the invalid questions were retained in the analyses of the question papers for comprability with other question papers.

147 The only data, other than the total mark for each paper that was common between this study and the WCED data for the entire cohort of candidates, was the total marks for each of the major questions on each question. paper. The WCED calculated the question means, as percentages, for all candidates who wrote each question paper but did not calculate the standard deviations. Unfortunately, the WCED were unable to supply the raw data for the entire cohort (T. Hamman, personal communication, July 19, 2010) so this author was unable to calculate means and standard deviations for the individual symbols of achievement or the standard deviations for the entire cohorts of WCED candidates. Both sets of statistics would have enriched a discussion about the representivity of this study's samples.

Table 5.4 A comparison on the percentage marks obtained by all candidates who wrote the 2005 and 2006 WCED SC Biology examinations, and by the samples of candidates selected for this study.

		2005							2006						
		N/n	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**All**	**N/n**	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**All**
HG Paper 1	WCED	8 452 [a]	69.2	51.0	54.1	42.4	39.0	53.4	8 683	64.0	54.2	64.5	58.2	56.3	60.0
	This study	1 001	59.8	44.3	44.2	36.3	32.2	45.4	975	55.5	46.1	58.1	48.4	45.1	51.3
			(23.8) [b]	(22.3)	(26.2)	(20.1)	(26.2)	(22.3)		(22.8)	(24.3)	(20.1)	(26.4)	(31.1)	(22.5)
HG Paper 2	WCED	8 452	54.8	57.0	64.4	39.0	42.5	51.9	8 683	56.2	42.8	40.9	40.6	46.7	46.8
	This study	1 004	48.1	50.2	57.1	34.9	35.1	45.5	975	48.3	35.4	33.9	34.3	40.1	39.7
			(23.0)	(22.3)	(22.5)	(24.7)	27.9	(22.1)		(23.2)	(23.7)	(25.0)	(22.6)	(22.7)	(21.9)
SG Paper 1	WCED	18 179	39.4	31.8	37.5	23.2	45.1	36.1	18 410	43.7	24.0	36.3	33.2	57.2	39.7
	This study	982	48.8	41.1	46.2	32.9	50.5	44.4	881	51.7	34.0	43.4	42.4	63.3	47.8
			(23.3)	(24.8)	(24.8)	(25.7)	(20.1)	(21.7)		(20.7)	(25.0)	(21.6)	(23.4)	(19.0)	(19.4)
SG Paper 2	WCED	18 179	39.0	22.0	14.2	24.8	22.1	26.9	18 410	29.7	25.6	28.9	28.1	19.3	26.9
	This study	854	52.2	34.4	21.9	38.3	34.6	38.9	881	38.1	33.7	36.8	36.4	28.6	35.3
			(20.4)	(21.0)	(15.1)	(22.9)	(21.5)	(21.5)		(20.1)	(20.5	(20.0)	(20.9)	(22.1)	(18.7)

Note:

a Each cell reports the mean mark for that group of candidates.

b The standard deviations are reported in brackets for this study but were not unavailable for the WCED.

and 5.2). Despite that the samples were selected differently in 2005 and 2006 (Section 5.2.2), the mean marks from all four HG samples were consistently lower than the mean marks for the corresponding entire cohort,[148] and the mean marks from all four SG samples were consistently higher than the mean marks for the corresponding entire cohort[149] which each represented. Any inferences made from the results of the analyses of candidates scripts about the populations of students who wrote the different examinations, will be noted accordingly in later sections of this thesis.

As the marks allocated by the markers at the time of marking were used in the script analyses in this study, it was necessary to test the reliability of the marks allocated. All examining bodies in South Africa have a practice whereby candidates who suspect that they have received an incorrect (low) mark can lodge an appeal to the examining body to have their script re-marked. This researcher had hoped to use the number of re-marked scripts which involved a changed or unchanged mark ,to determine the reliability of the script marks. Unfortunately, the WCED were not able to supply the necessary records about Biology SC re-marks (T. Hamman & J. Parboo, personal communication, July 26, 2010). The author recognizes that observations relating to the performance of the students might change if the reliability of the candidates' marks could be improved but this criterion would have involved the re-marking the 7 553 scripts with multiple markers to obtain inter-rater reliability measures and a better informed consensus on the marks awarded, a project for which there were no resources. Instead, once the script data were captured electronically they were checked for obvious discrepancies by checking for over-maximum marks for each of the questions. The discrepancies detected, and corrected, are noted below since they have the potential to improve both the quality of the SC Biology examination marking processes in the future and future research into subject level performances on the SC examinations.

Approximately 52% of all the errors recorded in the marks allocated to candidates scripts in the samples were the result of data capture errors (in this study), and the remainder were the result of mismarking or incorrect subtotals, at the time of marking (WCED). Most of the data capture errors were the results of candidates incorrectly numbering questions, leaving out questions, or answering questions out of sequence, yet where markers were able to recognize the questions the questions

[148] The distributions of scores in HG populations were skewed towards the symbols A to F, with fewer candidates achieving the symbols FF to G.

[149] The distributions of scores in SG populations were skewed towards the symbols E to H, with fewer candidates achieving the symbols A to D.

were marked. Data capturers focused on question numbers and the mark recorded for each. They could therefore not have be expected to detect incorrect marks allocations because in most case markers did not change the question numbers accordingly. Some markers did not provide sub-totals for sub-questions and other markers were not always vigilant or consistent as to where on the scripts they placed the sub-totals. Such inconsistencies meant that the data capturers had to try to work out how marks were allocated. In cases where a question carried maximum marks in the memorandum, some markers awarded more marks than they should have.[150] When incorrect factual answers that had been marked correct were detected, the marker's mark was used based on an assumption that this error would have consistently occurred.[151]

5.2 **Source analysis**

The analysis of question papers involves the descriptions and analysis of those variables which might be expected to influence candidate performance, the validity of an examination and the construction of the meaning of students' performance in these examination. The variables for this study that were identified in Chapter 3 are descriptors associated with the structure of an examination question paper; the structure of questions and the structures of answer that make up that question paper; and the topic(s) and cognitive demand (performance expectations) of each question, which together describe the content standards of an examination. The content standards together with the structures of the question paper, the questions and the answers were used to generate a unique profiles[152] for each question paper.

Data from candidates' answer scripts were then extracted according to the scorable events (Section 5.2.1) identified in the profile generated for the corresponding examination question paper. Collectively, the pre-defined symbols, A to H and their cut-scores, together with the content standards, and the structural aspects of a question paper, explain the performance standards of a specified group of students in that examination.

150 For example, a question could have had a maximum mark allocation of four marks and the memorandum indicated that there could be eight possible marks. In this instance, a candidate who achieved all eight possible marks would only receive four marks for this question.

151 The marking of scripts is moderated at each provincial marking centre, and centrally by Umalusi (Chapter 2).

152 A question paper could consist of one profile only, that is if the question paper has no choice of questions. Where there is a choice of questions, if the structural aspects and/or the content standards of the question paper differ between the choices, more than one unique profile will be generated.

5.2.1 Examination question papers

The analysis of the examination question papers involved the scrutiny of the complete examination question paper as a whole and the examination of individual or parts of individual questions.[153] Descriptions of individual questions required the author to work through to a solution to each question when official memoranda were not available.[154] The variables which resulted from this analysis of a complete examination question paper (Section 5.2.1.1) and an item-by-item analysis of the questions on each examination question paper (Section 5.2.1.2) together build the profile(s) of each question paper (Section 5.2.1.3). The profile of a question paper informs the way in which data are collected from the 2005 and 2006 candidates' answer scripts because identified in the profile are the individual scorable events for which candidates' marks are required.

Biology as a scientific discourse combines natural or everyday language with academic language, scientific terms, mathematical terms, symbols, graphs, diagrams and tables and therefore the SC Biology examinations should reflect those elements. Therefore, the variables used to collect data about the SC Biology question papers and candidates' answer scripts in this study were selected to be representative of these different facets of Biology. Measures of language usage in the question papers are not as detailed as those suggested by Abedi (2006) to measure the linguistic complexity of test items, because language is not a focus of this study and because the breadth of this study does not admit its incorporation.

The unit of analysis used for extracting and coding the data is a *scorable event* (Britton & Raizen, 1996).[155] Scorable events are "the smallest [discrete] questions in an examination that cannot be broken down into more sub-questions" (Britton & Raizen, 1996, p 269).[156] Each scorable event is weighted according to the mark(s) that it carries. For example, a question with four sub-questions worth 2, 3, 4 and 5 marks each, was considered as at least four separate scorable events, each weighted by the corresponding marks. An example of how scorable events were recognized is given in Chapter 4 (Appendix 4.4).

153 "[Q]uestion" as used in this study includes a "test item" or an "instruction" that "requires a student response under certain conditions and specific scoring rules" (Haladyna, 1997, p. 36).

154 Prior to 2001, official memoranda were not released to the public by examining bodies.

155 Edwards and Dall 'Alba (1981, p. 162) called this process "uniti[z]ing".

156 Scorable events are different from the units of analysis, called "test tasks", used by Valverde (2005, p.33) because a test task could be simultaneously located in multiple categories of content (topic) or multiple categories of performance expectations. Here scorable events are each located in a single topic and a single performance expectation.

In this study the total number of marks allocated to a question papers varied, (e.g., 200, 300, 400) within and between years. Aitchison (1986) advocated transforming the total marks of question papers when they varied, seeking comparability. This study used untransformed marks throughout the analyses.[157] This decision arose from the view that the length of a question paper (in hours) is contributes towards the difficulty of a question paper (Britton et al., 1996a). Many question papers used in this study were written in the same period of time, (i.e., three hours) and yet carry a different number of total marks. This variation meant that a candidate writing a question paper of 400 marks in three hours of examination is potentially[158] under more pressure (i.e., might find the question paper more difficult) than a candidate writing a question paper of 300 marks in the same time. In order to report on and examine these potential differences, the untransformed data was used. When comparability was required, the marks were represented as proportions of the total paper mark.

5.2.1.1 *Complete examination question papers* (Appendix 5.2 [coding form])

The following variables, modified from Britton (1996), concerning each examination question paper and its memorandum (examination question papers from 2001 to 2007 only) were extracted from each entire examination question paper:

(i) The year of the examination
(ii) The examining body
(iii) The number of question papers comprising the examination
(iv) The total time candidates are allowed to spend on the examination
(v) The total mark of the examination
(vi) The total marks allocated to each of the question types that appears in each of Sections A, B and C (multiple choice questions [MCQs], matching columns or diagram labels, item/statement, terminology/one word, missing words/labels, identifying incorrect/correct labels, short questions) [159]
(vii) The number of options in MCQs

157 Some question papers weighted different sections of their question papers. Where this weighting was indicated on the question paper the weighted marks were used. The implications of weighting questions at the time of paper-setting is discussed in Chapter 6.

158 The word 'potentially' is used because the 'value' of a mark can vary within and between similar questions that appear on the same question paper. This is discussed in Chapter 6.

159 Short questions are those questions in Section A that cannot be characterized by any of the other categories. These questions may be similar to the short questions of Section B and Section C.

(viii) The section of the examination/paper that offers a choice of questions, if present, and the type of choices offered

In the initial proposal for this study it was intended that a variable called 'source of challenge' (Webb, 2007) be captured. This variable was to have recorded indicators of invalid questions, for example, those questions that used irrelevant text or diagrams being irrelevant to the answer or difficult language. A number of such invalid questions was observed in the question papers from 1994 to 2000 but, in the absence of official memoranda, there was no record of whether these questions were detected as being invalid at the time of writing or how the marks might have been adjusted. For question papers from 2001 to 2007 the memoranda indicated that all candidates were given full marks for invalid questions. These invalid questions were noted and were used in the analysis of question papers but not in the anlyses of candidates' answer scripts, except where noted.

5.2.1.2 *Item-by-item categorization* (Appendix 5.2 [coding form])

The scorable events in each examination question paper were identified. A total of 11 006 scorable events were identified and the number of scorable events varied per question paper. A summary of the number of scorable events which were identified for each of the question papers is given in Appendix 5.3. For each scorable event, on each examination question paper, the following variables, modified from Britton and Raizen (1996), were recorded :

(i) The scorable event label (question number, sub-question number or if a question/sub-question is broken down into a number of scorable events each with a unique identifying number.

(ii) Mark – the maximum mark awarded to each scorable event in the question paper. It is used to weight the relevant variables (iii) to (ix) described below.

(iii) Type of answer required by question.

   a. *Choose correct answer* [160]– requires candidates to choose one answer from a given subset. This type includes an answer of yes/no, true/false.

   b. *Free response answer* [161]– requires candidates to construct his/her own response.

---

[160] Also known as selected response items (Linn, 2006; Downing, 2006).

[161] Also know as constructed response items (Welch, 2006).

(iv) Length of text, including the stem,[162] in the question needed to answer question.

a. *One sentence*

b. *Two to three sentences*

c. *More than three sentences*

For a scorable event where a paragraph was given and candidates were expected to fill in missing words, each missing word was considered to come from one sentence because each answer could have been answered by reading the single sentence in which it was located. To have considered such an item as a question of length more than three sentences might have implied comprehension skills not required by the question. An example of this kind of question was Question 1.5 in the National 2002 HG Paper 2.

For scorable events, such as MCQs, the generic instructions about how to present the answers were excluded from these analyses.

Ebel (1965, quoted by Board & Whitney, 1972, p. 225) labelled material included in the stem of a question but which was not necessary to answer the question as "window-dressing". Even though the window-dressing in a questions might be unrelated to the answer required by the question, it is often used to make the scenario posed in the question more realistic to students answering the questions (Haladyna, 1997). Despite such a possible advantage of window dressing, Haladyna (1997) and Haladyna, Downing and Rodriguez (2002) recommended that the stem of multiple choice items be kept as brief as possible. A number of the multiple choice questions and the free-response items analyzed in this study included window dressing. This author decided to include window dressing in the length of scorable events because students would have had to read the window dressing of a question before they could determine that the material was not necessary to answer the question.

(v) Use of non-text elements as part of the question text for particular scorable events

a. *Diagrams* (includes flowcharts, photographs and micrographs)

b. *Graphs*

c. *Tables.*

[162] "[T]he stem is the part of the item that asks the question, sets the task the student must perform, or states the problem the student must solve" (Nitko & Brookhart, 2007, p. 148).

(vi) Length of text required to answer (free response answers only)

a. *One- or two-term answers* requiring only one or two words or terms, or requiring a value to be read directly from a graph

b. *Short answers* requiring one to three sentences of text, or quantitative answers requiring only one formula or equation in a single step, or giving an equation

c. *Extended answers* requiring four or more sentences of text, or quantitative answers requiring multiple-calculations and/or more than one formula or equation or reading from a graph and making a calculation or a decision.

Where a scorable event required a particular number of stand-alone answers, each answer was considered separately. To have considered this kind of answer as one entity might have implied coherence to an answer that required no coherence or particular order.

(vii) Production of non-text elements required for the answer (free response answers only)

a. *Diagrams* (includes flowcharts and crossword puzzles)

b. *Graphs*

c. *Tables.*

(viii) Topic – selected from the topics as stipulated in the CBS and the NGDB as summarized in the 12 topics, and at category outside of the syllabus, listed Table 5.5. This table was compiled using topic headings listed in the appropriate core syllabuses and guideline documents, DNE (1984a, 1984b) and DoE (2002b).

Where a sub-topic, (e.g., diseases associated with malnutrition and some of aspects of human homeostasis) appeared in two different topics, a scorable event was coded using the topic in which that sub-topic was first given in the CBS, to ensure consistency in coding.[163]

When knowledge from two different topics was required for one scorable event, the marks were divided proportionally between the two topics.

Knowledge required and learned prior to Grade 12, according to the Grades 10 and 11[164] CBS, was treated as being within the relevant topic in Grade 12.

163 This device means that for the national examinations from 2001 to 2007 the topic totals calculated in this study may differ slightly from the topic totals required by the policy introduced for examinations in 2001 (see Chapter 2).

Table 5.5 A list of topics examined in the Biology SC Examinations 1994 to 2007 as grouped in the CBS (DNE, 1984a, 1984b).

Code [a]	Grouping [b]	Topics
11	1	Biological compounds
12	1	Enzymes and co-enzymes
21	8	Angiosperm physiology: water relations
22	7	Angiosperm physiology: growth & development
23	2	Angiosperm physiology: photosynthesis
30	3	Cellular respiration
41	4	Aspects of human anatomy and physiology: nutrition
42	5	Aspects of human anatomy and physiology: gaseous exchange
43	9	Aspects of human anatomy and physiology: excretion
44	10	Aspects of human anatomy and physiology: co-ordination
45	12	Aspects of human anatomy and physiology: circulation [c]
41/51	4	Homeostasis: human nutrition
42/52	5	Homeostasis: gaseous exchange
43/53	9	Homeostasis: excretion
44/54	10	Homeostasis: co-ordination
55	11	Homeostasis: thermoregulation & tissue fluid
61	6	Population dynamics [d]
71	9	*Amoeba* and earthworm osmoregulation [e]
81	13	Outside of syllabus

Note:
a Topic grouping used in this study. There is no significance to the ordering. The topics are arranged in this order in the DoE policy documents for 2001 to 2007.
b Topic grouping from DoE (2001a) used in analyses.
c IEB only, 1996 to 2003.
d Not used by IEB 1996 to 2003.
e Removed from examination requirements from 2003

(ix) Performance expectation – what a candidate is required to demonstrate when answering the scorable event was classified using the PET developed in Chapter 4 (Figure 4.6). The PET groupings are Memorize, Perform Procedures, Explain, Analyze and Apply. The PET used the CBS and the NGDB (DNE, 1984a, 1984b; DoE, 2002b) as a reference point when determining the familiarity of the context of a performance expectation.

[164] Called Standards 8 and 9 in 1994 and 1995.

Variables generated for each scorable event

(x) *Content* – a combination of (viii) and (ix) described above, comprising a total of 65 possible content categories (13 possible topics multiplied by 5 possible performance expectations). Each combination of topic and PET level which in this study defined a 'content' category, was called a 'signature' by Valverde (2005, p. 33),[165] or COMB in Appendix 5.2 and COMB 1 to COMB 65 in Appendix 5.4.

5.2.1.3 *Profile of complete examination question papers* (Appendix 5.4[coding form])

Each examination question paper was profiled in terms of its structure (of the question paper as a whole, of the questions and of the answers) and its content (topics and performance expectations). As a result, each question paper had its own unique profile(s) which enabled comparisons of examination question papers to be made within and between years. Where there was a choice of question(s), the question paper could have had more than one profile. Multiple profiles resulted when the choices offered within a question were not exactly the same, that is, they differed in one of the variables described in Section 5.2.1.1 (v) and (vi), or (iii) to (x) described above. The specific questions which comprised each of the 221 unique profiles identified for each of the question papers are given in Appendix 5.5. Different combinations of the profiles of a question paper with multiple profiles were used in different analyses, depending on the particular variables being described for each question paper. For example, the IEB 1997 question paper had eight unique profiles, if all the variables which comprise the entire profile for a question paper are considered. If these eight profiles were grouped according to the question types or according to the free response answer types, four unique combinations of the profiles emerge (1 and 2; 3 and 4; 5 and 6; 7 and 8). If these profiles were grouped with respect to any aspects of content, that is, topics and performance expectations, four different unique combinations of the profiles emerge (1 and 3; 2 and 4; 5 and 7; 6 and 8).

Other than the year, the examining body, the number of the paper in the examination (i.e., 1 = 1 paper only, or 1 of 2 papers), and the profile number for the question paper (= 1 if there was no choice of questions in the paper or if choices resulted in identical profiles), each question paper had a value for each of the variables described in Section 5.2.1.1 (v) and (vi); and (iii) to (x) described in this section. The value for each variable in a question paper was obtained by summing the mark allocation data for that variable from all the scorable events.

165 Valverde used 'content' to describe subject matter topics (Valverde, 2005, p. 33).

Two variables, that is, the indices derived for breadth and depth of an examination were excluded from the profiles since they were derived from measures of content already represented in the profile. These two new variables were generated from totals within a question paper for the variables, topic and PET category, described in this section ([viii] and [ix]), for use in the comparison of question papers.

Variables generated for each question paper

(i) *Breadth of knowledge* (BOK) – reflects the number of topics covered and the relative emphasis on each in a question paper. It was measured by index obtained by using the following equation:

$$BOK = \frac{T(T-1)}{\sum t(t-1)}$$

where T = total number of marks for the complete question paper
and t = the total number of marks for each topic.

The mathematics of this equation is borrowed from ecological studies, the Simpson diversity index, which is used to compute the *relative abundance* of species in a habitat (Allott, 2001; Indge, 1997). In this study the equation weights the identified topics according to the number of marks each carries on a question paper. Therefore in this study, BOK represents the *relative emphases* of different topics. Alignment studies (see Chapter 3) and Tamir (1996) offered other ways to measure breadth of cover in assessment but none of these methods capture the breadth of cover in the way that this equation does.

Figure 5.2 shows the use of this index on four hypothetical question papers exhibiting differing total and different emphasis on different topics. Amongst the hypothetical question papers, the papers with the widest range in the emphases of different topics (Examination 1) has the lowest BOK value and the highest BOK values are found for the question papers with all topics covered equally (Examinations 2 and 4). Examination 3 covers all the topics equally but it covers fewer topics. The BOK index involves the number of topics. The BOK index for a 200 mark question paper split evenly amongst 5, 10 or 20 topics are 199/39, 199/19 and 199/9, approximately 5, 10 and 21, If the marks are not evenly spread amongst the topics, the BOK diminishes, and may be a number smaller than the actual number of topics. The number of topics actually covered by a question paper should therefore be considered when interpreting a BOK index. In essence, the BOK counts an effective number of notionally equally allocated topics (with topic marks greater than 1) in place of the notional number of topics present. The BOK also treats each actual topic as equivalent to any other topic, in effect as if an ideal content has all its topics evenly covered, and hence may have limitations and may be contested.

Breadth of knowledge = BOK

$$\text{BOK} = \frac{T(T-1)}{\times\, t(t-1)}$$

Where T = total number of marks for the complete question paper and t = the total number of marks for each topic. For example, using the data below:

	Number of marks for each topic			
Topic	**Examination 1**	**Examination 2**	**Examination 3**	**Examination 4**
Biochemistry	20	60	100	80
Respiration	10	60	100	80
Photosynthesis	10	60	100	80
Human sense organs	100	60	0	80
Human excretion	160	60	100	80
Total	300	300	400	400

$$\text{BOK Examination 1} = \frac{(300 \text{ x } 299)}{(20 \text{ x } 19) + (10 \text{ x } 9) + (10 \text{ x } 9) + (100 \text{ x } 99) + (160 \text{ x } 159)}$$

$$= \frac{89\ 700}{35\ 900}$$

$$= 2.50$$

$$\text{BOK Examination 2} = \frac{89\ 700}{17\ 700}$$

$$= 5.06$$

$$\text{BOK Examination 3} = \frac{159\ 600}{39\ 600}$$

$$= 4.03$$

$$\text{BOK Examination 4} = \frac{159\ 600}{31\ 600}$$

$$= 5.05$$

Figure 5.2 Examples of calculations using the BOK index.

(ii) *Depth of knowledge* (DOK) – is a measure of the cognitive demand[166] of a question paper defined as the ratio of marks allocated to HOCS[167] and to LOCS[35] per examination paper, obtained by using the five PET categories (Memorize, Perform-Routine-Procedures, Explain, Analyze and Apply) as:

$$\text{DOK} = \frac{\text{(total marks Analyze + total marks Apply)}}{\text{(total marks Memorize + total marks Perform-Routine-Procedure + total marks Explain)}}$$

The DOK is a natural way of comparing complementary percentages, and has an equivalent in the notion of odds. In this study DOK is used to indicate the emphasis on HOCS in a question paper, and to compare relative emphases of HOCS between question papers.

Each question paper was profiled twice, with three days between each coding, to ensure consistency between the coding of all the question papers. The data collected for each profile was collated into Microsoft Office Excel 2003 spreadsheets using the format shown in Appendix 5.4.

Some studies have measured the DOK of assessments differently but none offered defensible reasons for their choice of method. For example, Freeman et al. (2011), and Haak et al. (2011) used what they call the Weighted Bloom's Index (WBI) as a measure of the cognitive demand of an examination question paper. The WBI is calculated giving each of the six BTEO categories a rank of 1 (Knowledge), 2 (Comprehension) up to 6 (Evaluation). Therefore the WBI assumes that for example, evaluation measured as a 6 is three times the magnitude of comprehension measured as a 2. Freeman et al. (2011) and Haak et al. (2011) offer no empirical support for ranking or weighting the BTEO categories in this way. For this reason the author did not use the WBI as a measure of the cognitive demand of examination question papers in this study.

5.2.2 Candidates' answer scripts

The analysis of the candidate answer scripts involved recording, for each student, the marks achieved for each of the scorable events identified for the corresponding question paper. The data collected from each student was captured directly into Microsoft Office Excel 2003 spreadsheets.

166 See Chapter 4 for the explanation of the relationship between cognitive demand and performance expectations.

167 See Chapter 4 for how LOCS and HOCS are used in this study.

5.2.3 Descriptive and analytical methods

In this study, the analyses of the data collected from examination question papers and student answer scripts used the process of data exploration (van Dantzig, 1978) with the intention of interactively detecting patterns, anomalies and relationships in the data which might not have been immediately obvious or anticipated. Initial analyses used univariate analyses to understand the absence, presence and nature of trends in the data over time (examination question papers) or patterns and contrasts between different sample populations of students (candidates' answer scripts). The stability of the relationship between the two variables, topic and performance expectation that constitute the content standards in this study, was examined between pairs question papers using a similarity index which is mathematically explained below. Subsequent multivariate analyses, that is, cluster analysis (locating sub-groups of similar examination question papers), discriminant function analysis, factor analysis and The mathematics of this equation is borrowed from ecological studies, the Simpson diversity index, which is used to compute the relative abundance of species in a Cronbach's alpha (applied to candidates' answer scripts) were used to further explore potential relationships between the variables which were not evident from the univariate analyses and the content analyses.

The patterns, relationships and anomalies which resulted from the kinds of analyses described briefly above were used to generate the ideas, relationships and inferences which partially inform answers to the research question and sub-questions, 2 to 7, in Chapter 7.

5.2.3.1 *Univariate analyses*

Initially the variables described for question papers (Section 5.2.1) and for candidate's answer scripts (Section 5.2.2) were analyzed and summarized separately using Microsoft Office Excel. The descriptive data generated by these analyses was presented through visual displays (i.e., tables, and graphical representations using Microsoft Office Excel 2003). Correlation coefficient analyses (Microscoft Office Excel 2003) were used to summarize the strength of relationships between the different variables within examination question papers.

5.2.3.2 *Comparison of content*

Tyler (1949) recommended that educational objectives, as indicators of what students should learn, should include both the knowledge (topics) and the specific behavior(s) that would indicate understanding or skill relative to that knowledge. Other researchers (e.g., Champagne, 1990); Hirsch, 1996; Knapp, 1992; Perkins, 1998b; Perkins, & Salomon, 1989) argued, by drawing on modern cognitive research, that higher order or critical thinking skills which have become so valued within education cannot be viewed independent of the discipline or knowledge to which they are inextricably linked. Similarly, Zoller and Tsaparlis (1997, p. 118) related student performance to both "the specific area of content [topic] being assessed and the generic cognitive abilities". Despite such recommendations in the literature many syllabuses around the world, including the core syllabuses on which this study is based (Chapter 2), continued to list topics and their content independent of a separate list of generic cognitive skills or performance expectations. In an USA alignment[168] study of the relationship between the content of student instruction and gains in student mathematics achievement, Gamoran, Porter, Smithson, and White (1997) found that using the intersection between content (which described the mathematical topics) and cognitive demand (which these authors called performance expectations) as the locus of a variable indicating student gains in achievement, gave higher statistical correlations than when topic or cognitive demand were used alone. A similar relationship, that is, an interaction between topic and cognitive level, was reported in another study of student mathematics achievement (Gierl,1997) and is consistent with the view that cognitive skills only have meaning when viewed together with content (Marzano & Costa, 1988). Porter (2002) argued that for alignment studies, in general, much would be lost by reducing research instruments to only either topic or performance, and by reducing studies to topics only or to cognitive demand only. Porter also advocated the use of this 'single language' (content)[169]for measuring content to ensure "description at a consistent level of depth and specificity" when comparing various components of the curriculum (Porter, 2002, p. 3). Conceptualizing content in this way embraces the working definition for content standards used in constructed in Chapter 3, namely, that content standards refer specifically to what students should know (topics) and be able to do with what they know (cognitive demand). This two dimensional 'language' for content has been developed to show graphically, using topographical maps, the content which is emphasized, or not emphasized, in USA alignment studies (e.g., Porter, 2001a, 2001b, 2002; Porter et al., 2009).

In Chapter 3, the author argued that in the absence of explicit content standards in a curriculum, descriptions of the content as conceptualized by Gamoran et al. (1997) can be used to explicate the

---

168 See Chapter 3 for discussion about alignment practices and how they inform this study.

169 Some authors, e.g. Liu, Zhang, Liang, Fulmer, Kim & Yuan (2008), Liu and Fulmer (2008) and Fulmer (2010), incorrectly quote Porter (2002) as using "content" to refer to what he termed "topics". By using this construction these authors did not acknowledge the subtlety of Porter's view of content as being more than topic only. Wineburg (1997) and Gross (2009) also advocated that content be used to refer to both substantive knowledge and cognitive processes.

implicit content standards from the examinations. The two-dimensional content 'language' is used to graphically compare the performance of candidates (2005 and 2006) with respect to content and to examine how the content standards generated for two selected groups of examining bodies changed between 1994 and 2007.

From the work of Gamoran et al. (1997), Porter (2002) developed an alignment index which could describe the match (or mismatch) between various curricular components. The interpretation of the value of the alignment index is not straightforward (Porter, 2002) but it does allow comparisons to made between different curricular components provided they use the same content framework. In this thesis the relationship between the two variables, topic and performance expectation, that constitute the content standards was examined for all pairs of the question papers using a similarity index, a description of which follows.

A *similarity index* (SI), which is mathematically the same as Porter's alignment index was used to indicate the similarity between the content of any two different examination question papers, as characterized by the proportions in their topic by demand data matrices. The equation below describes the SI as used in this study to compare the content of one question paper with that of another question paper.

$$SI = 1 - \frac{\sum |X - Y|}{2}$$

where X = the cell proportion[170] in a cell in one matrix and
Y = the corresponding cell proportion in the other matrix,
and summation is made over all cell positions in the data matrices

Conceptually this SI can be thought of as "the sum of cell-by-cell" intersects (Porter, 2002, p. 5) and theoretically have a value of 0 to 1.0 (Porter, 2002, p.5). The larger the SI, value the greater the similarity between any two question papers being compared.

The use of this similarity index using three hypothetical examination question papers with different content emphases is shown in Figure 5.3. In the hypothetical examinations, in terms of their content,

170 Proportions of the total marks are used to facilitate comparisons between question papers with different total marks (e.g., HG versus SG).

Examination 1 is less similar to Examination 2 (SI = 0.40), than it is to Examination 3 (SI = 0.90), as confirmed by inspection of the three examinations.

Similarity index = SI

$$SI = 1 - \frac{\sum |X - Y|}{2}$$

Where X = the cell proportion (of content expressed as a proportion of the total examination) and Y = the corresponding cell proportion (of the same content expressed as a proportion of the total examination), and summation is made over all cell proportions in the data matrices

Performance expectations

Topics

Examination 1

.1	0	.1
0	.3	.1
.2	.2	0
0	0	0

Examination 2

.4	0	0
0	.2	0
.1	.2	.0
0	0	.1

Examination 3

.1	0	.1
0	.2	.2
.2	.2	0
0	0	0

Comparing examination 1 with examination 2:

$$SI\ (1,2) = 1 - \frac{|.1 - .4| + |.1 - 0| + |.3 - .2| + |.1 - 0| + |.2 - .1| + |.2 - .2| + |0 - .1|}{2}$$

$$= 1 - \frac{(.3 + .1 + .1 + .1 + .1 + 0 + .1)}{2}$$

$$= \mathbf{0.40}$$

Comparing examination 1 with examination 3:

$$SI\ (1,3) = 1 - \frac{|.1 - .1| + |.1 - .1| + |.3 - .2| + |.1 - .2| + |.2 - .2| + |.2 - .2|}{2}$$

$$= 1 - \frac{(0 + 0 + .1 + .1 + 0 + 0)}{2}$$

$$= \mathbf{0.90}$$

Figure 5.3 Example of calculations using the Similarity Index (SI).

5.2.3.3 *Multivariate analyses*

> Multivariate analysis provides statistical methods for study of the joint relationships of variables in data that [incorporate] intercorrelations. Because several variables can be considered simultaneously, interpretations can be made [about associations between variables] that are not possible with univariate statistics. (James, 1990, p. 129)

Due to the exploratory nature of this study of examination question papers and candidates answer scripts different, multivariate techniques were used. Cluster analysis (using question papers), discriminant function analysis and factor analysis, and Cronbach's alpha reports (using candidates' answer scripts) were used to further explore potential relationships between the variables which were not evident from the univariate and bivariate analyses.

Cluster analysis was used in this study to explore potential groupings of different but similar question papers (within and between years) based on combinations of the variables that characterize each question paper (the content standards). Cognizant that competence (or mastery) "is a complex and multifaceted proficiency" (Glas, 2003, p. 76), that is difficult to capture by unidimensional scoring (Glas, 2003) two other multivariate methods, that is, discriminant function analysis and factor analysis, were used to analyze the scores from candidates' answer scripts. The purpose was to describe and understand the potential relationships between different student performances (the performance standards) and the role that cut-scores play in differentiating between different student performances.

The software package used for the multivariate analyses was Statistica 9 for 2010 (STAT9201007). The multivariate methods and the reasons why they were used in this study are briefly described below.

a. ***Cluster analysis***

Cluster analysis is a descriptive method that classifies a set of individual items under investigation into groups based on their internal similarity (Bartholomew, Steele, Moustaki & Galbraith, 2002). A "[p]roblem is that cluster analysis can produce clusters whether or not natural groupings exist, and the results depend on both the similarity measure chosen and the algorithm used for clustering"

(James & McCulloch , 1990, p. 148). Asking what it is that members of a group may or may not have in common would help to define any groups identified numerically(Bartholomew et al., 2002).

Cluster analysis was used in this study to detect patterns, if any, of similarity between the different question papers within and between years. The unique profiles of each of the question papers were used in the cluster analyses. Using this method,[171] question papers were grouped according how similar they were with respect to their profiles. Groups of question papers identified by the cluster analyses as being similar or less similar were then examined. By examining the univariate data, for possible descriptions were sought for common features specific to a cluster.

b. ***Internal consistency of the examinations***

Internal consistency of test marks "is the most common type of reliability since it can be estimated from giving one form of test once" (McMillan & Schumacher, 2001, p. 246) – the SC Biology examinations are such tests. Therefore, it was important to determine how consistent the marks that students achieved in each SC Biology examination were. Deciding on which measure of internal consistency to use was problematic because each examination included a mixture of different kinds of questions, each best suited to being tested for internal consistency by a different method. In split-half reliability measures the items of a test are divided into comparable halves, of similar difficulty and the correlation coefficient for pairs of halves is calculated (McMillan & Schumacher, 2001). In the SC examinations analyzed in this study, dividing the items into two appropriate halves could not be done because of the combination of factors which influence difficulty (see Section 5.2.3.4). The Kuder-Richardson formula can be used on a single administration of a test, and is used when the answers to the items comprising the test are either right or wrong (McMillan & Schumacher, 2001). Varying proportions of items comprising the SC examinations investigated in this study had binary (0/1) or right or wrong answers,[172] but more questions had a range of answers, and so the Kuder-Richardson method was considered to be unsuitable. The Cronbach's alpha is considered best suited to questions where there is a range of possible answers (McMillan & Schumacher, 2001), and a range of different kinds of questions (Nitko & Brookhart, 2007). Because each of the SC Biology examinations included a range of different kinds of questions (e.g., MCQs, short open questions and

[171] The algorithm used to cluster the question papers used unweighted, pair-group average and Euclidean distance.

[172] Examples from the SC Biology examinations with either right or wrong answers would be multiple-choice questions, and answers which carry only one mark per test item.

essay questions, Cronbach's alpha was used to examine the internal consistency of the scorable event data within the candidates' answer scripts.

c. ***Discriminant function analysis***

In the SC examinations Biology candidates were assigned a symbol (A to H) based on the aggregate mark which they received. This process meant that individuals with similar aggregate marks were grouped together and received the same symbol. The cut-scores which separated one symbol from another, and a separate a pass mark from a fail mark, remained the same each year.[173] As students with the same symbol may, or may not, have obtained their aggregate mark by successfully answering the same suite of scorable events or questions, this study sought to determine if, on the basis of the individual subscores, individuals were correctly classified within the outcome symbol. That is, did the aggregate scores classify any candidates differently than if other relevant combinations of achieved sub-scores were used? Since the aggregate mark was used to define the groups *a priori* for each analysis, the aggregate mark was not included in the discriminate functions analyses.

Data used for the discriminate functions analyses were the individual marks that each candidate obtained for the scorable events in the question paper that they had written.

d. ***Factor analysis***

Exploratory factor analysis was used to investigate the nature of the relationships, if any, between the marks that students achieved for different scorable events because "it emphasizes the analysis of relationships among attributes [marks for scorable events]" (James & McCulloch, 1990, p. 133). These relationships might indicate the possible presence of a trait, or latent variable (Everitt & Dunn, 1991), which was not or could not be measured in a study. In this study, exploratory factor analysis was used to provide evidence that the set of scorable events in each examination did, or did not, measure the one general proficiency the examinations items were collectively designed to measure.[174] It was also used to determine if there were relationships between specific scorable events that had not been detected in the univariate and bivariate analyses.

[173] Cut-scores were the same for all subjects.
[174] Candidates received a single mark which signified their level of proficiency in Biology.

Data used for the discriminate function analyses were the individual marks that each candidate obtained for the scorable events in the question paper that they had written.

5.2.3.4 *Difficulty of examinations*

Scrutinizing the items comprising tests is useful in determining if some of these items put groups of test-takers at a disadvantage. That is, particular items may, if they are unfairly difficult or easy, introduce construct-irrelevance for some test takers and would therefore be unfair test items (Linn, 2006). The difficulty of curricula, especially examinations, based on the difficulty of the various items which comprise the test, have been undertaken in South Africa (Umalusi, 2009a, 2009b, 2010). In the Umalusi studies, subject experts rated the difficulty of each examination item "based on the perceptions and experience of the evaluators" (Umalusi, 2009b, p. 54) rather than on the students' actual responses to each examination item. Another method used to determine the difficulty of test items uses item analysis or differential item functioning (DIF) analyses (Livingston, 2006). Item analyses, such as DIF, can provide information about the difficulty and the discriminatory power of each of the items comprising a test (Livingston, 2006). Item analyses done after a test has been marked can give an indication of the full range of difficulty of the items on a test (Livingston, 2006). There are a number of complicated statistical methods used in item analysis which are described in Downing and Haladyna (2006). Livingston (2006, p.431) stated that the "simplest and most obvious measure of difficulty of an item for a group of test takers is their average score on the item", but cautioned that this measure could be misleading if the group of test takers was unusually strong or weak. Fitzpatrick, Ercikan and Yen (1998) successfully used this method in a study of rater consistency between years. Others, (e.g., Martiniello, 2009) used the proportion of students who answered an item correctly in the whole sample of students as an indicator of difficulty. Differential item functioning is used to determine discrepancies in the difficulty of test items for members of two or more groups of test takers with different backgrounds (Linn, 2006). For example, different methods of DIF analyses (Livingston, 2006; Welch, 2006) have been successfully used for comparisons between the test performance of different ethnic groups in the USA (Zieky, 2006).

Neither the Umalusi method previously described nor, the more sophisticated item analyses or DIF method of determining test question difficulty described in Downing and Haladyna (2006) were used in this study. The Umalusi method was not used because analyses of student scripts were excluded in the researchers' decisions about the difficulty of questions.[175] The Umalusi researchers also reported a lack of consensus, especially about the use of the category that they called moderately easy. Impara and Plake (1998) suggested that examiners are not good at identifying the differential task difficulty of items comprising assessments without the use of statistical methods. Alberts (2001) found that while teachers' estimates of the difficulty of examinations were consistent, they were only moderately correlated with pupils' results. Britton et al. (1996a) found that consensus between raters using a 3-point difficulty scale to rate questions was unreliable. Analyses such as DIF analyses could not be used in this study because the sampling method that was used to select the candidates' answer scripts[176] was not amenable to the types of statistical analyses required by this method.

Therefore, in this study the simplest of methods described by Livingston (2006) was used to measure the difficulty of each scorable event. That is the average score for each scorable event was calculated, as a percentage, for each of the eight 2005 and 2006 examination question papers, to analyze possible patterns in the difficulty of scorable events as reflected in by the marks achieved by the candidates in the samples. Because the samples for this study were selected from the full range of possible performance levels, Livingston's (2006) caution regarding sampling is unfounded. The calculated difficulty for each scorable event was then plotted against topics, performance expectations, and the structure of questions and answers to explore possible relationships between these variables and the difficulty of questions.

Obviously, a combination of the difficulty levels of each of the scorable events comprising an examination gives some indication of the difficulty of the examination as a whole. Britton et al. (1996a) considered a single rating for the difficulty of an examination to be unwise without considering a number of other factors which are linked to examinations. For example, Eubanks and Eubanks (1996) identified consideration of candidates prior coursework, or how they were taught, as being important in influencing whether questions constitute recall or analytical type questions (i.e., the difficulty of an examination). In addition, the number of different examinations a student must take in any one examination season and the manner in which examinations are administered and scored both influence the difficulty of examinations (Britton et al., 1996a). Other aspects of difficulty might arise because of specific context of an examination system (Britton et al., 1996b;

175 Discussions that this author has had over a number of years across the country with experts, like examiners and teachers, have indicated little consensus as to the type of questions that are difficult for students where there was no empirical data from student scripts to support the experts' opinions.

176 Samples of scripts were representative of each category of performance on each examination but were not collectively representative of the entire cohort of students who sat the examination (Section 5.1).

Schwille, 1996). More research is needed into what influences the difficulty of science assessments, especially given that the diversity of students is changing in many testing contexts (Penfield & Lee, 2010).

This study combined features of the South African SC Biology examinations with comparable work which described the difficulty of mathematics and science examinations in a number of countries (Britton et al., 1996). The South African data, which were derived entirely from the analysis of examination question papers and the policies which guided their setting and administration,, complement the descriptive analyses of the differences in the performance of candidates in different achievement categories (symbols) when the difficulty of examinations is presented and discussed in Chapter 6.

5.3 Chapter summary

This chapter completes the operationalization of the conceptual framework developed in Chapter 3 and attention to research sub-question 1. The first part of this chapter describes the data sources, that is, the 111 SC Biology examination question papers (HG 1994-2007 and SG 2001-2007), and the candidates' answers scripts for two of these years, 2005 and 2006 (HG and SG). The selection of the 7 553 candidates' answer scripts and their representivity of the population of students from which they were drawn was discussed. A section of this chapter described how the structural aspects of each examination question paper were studied, and how the content standards implicit in each question are explicated in the next chapter, Chapter 6. Scorable events are the units of analyses used in this study, and 11 006 scorable events were identified in this study. Combining some specific structural aspects of a question paper and its content standards results in the identification of 221 unique profile(s) for all the examination question papers. Each of the relevant profiles was then used together with the fixed symbols and their cut-scores to unpack the performance standards of the candidates' answer scripts for 2005 and 2006. The last section of this chapter explains two ways in which the comparative difficulty of the SC examinations will be determined, that is, by examination of the question papers as a whole, and by analyses of student performance. The results emanating from these three sections of research will be presented and discussed in Chapter 6 and will be further discussed in Chapter 7 in the light of the research question and sub-research questions (Chapter 1).

CHAPTER 6

ANALYSES OF QUESTION PAPERS AND CANDIDATES' ANSWER SCRIPTS

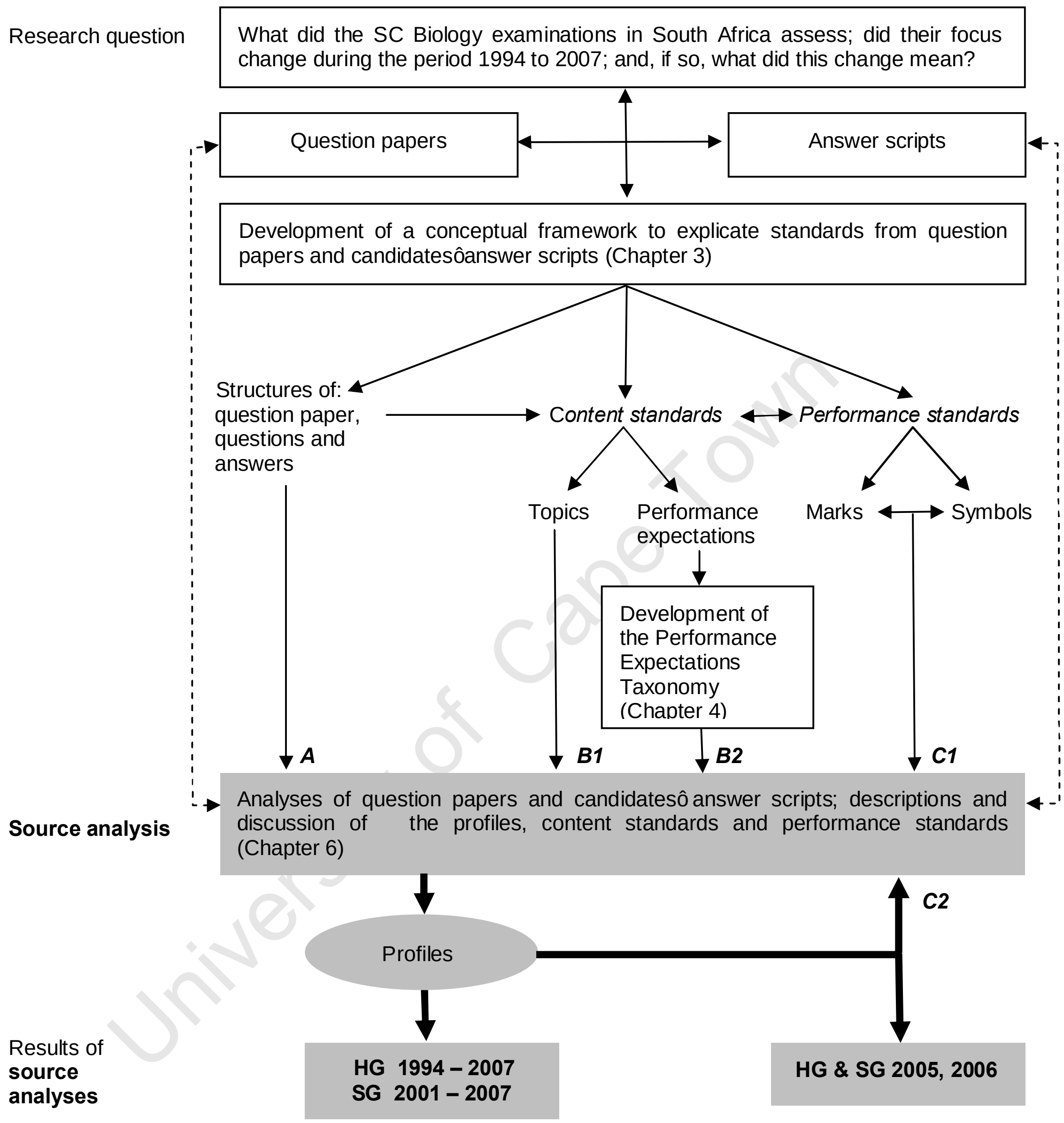


Figure 6.1 Relationship between Chapter 6 and the preceding chapters.

This chapter describes the analyses of SC Biology examination question papers and answer scripts in terms of standards as conceptualized in Chapters 3, and the operationalization of the conceptual framework in Chapters 4 and 5. Particular analyses presented in this chapter generate findings which potentially have a profound effect on South African policies and examination practices, but are outside of the research question and sub-research questions which will be specifically addressed in Chapter 7. For this reason they are intentionally discussed.

The conceptual framework developed to guide this study identified three strands of validity evidence that are important in understanding, and comparing, student performance in the SC Biology examinations. These strands are the *structural strand* (Figure 5.1[*A*]); the *content strand* (generates content standards) (Figure 5.1 [*B1* and *B2*]); and the *performance strand* (generates performance standards) (Figure 5.1 [*C1* and *C2*]). The relationship between the content standards and performance standards, as defined in Chapter 3, predicates that performance standards derive their meaning only when viewed together with the content standards. This claim means that the content standards of a question paper must be generated before the performance standards. The structure of question papers, questions and answers influence the difficulty of a question paper and, therefore, the corresponding performance of students. It is thus the *interplay* between the content, performance and structural strands that permit the attribution of meaning to candidates' marks in an examination, and these three strands *collectively* describe the standards of an examination.

In Chapter 4 the development of a tool, the PET, to determine the cognitive demand component of examinations was described. Chapter 5 was concerned with explaining how the content standards and the performance standards conceptualized in Chapter 3 could be explicated by analyses of SC Biology examination question papers and student answer scripts together with the structural aspects of examination question papers influence the content standards and the performance standards of examinations. A combination of data about the structural aspects of each examination question paper together with its content standards, derived using the methodologies described in Chapter 5, resulted in a unique profile, or profiles if the question paper offered a choice of questions. These profiles are used to compare the different Biology SC Biology examination question papers and to generate the performance standards from candidates who wrote the 2005 and 2006 WCED SC Biology examinations.

This chapter is divided into five main sections; the first section deals with the analyses of the examination question papers over the fourteen years, between 1994 and 2007. Descriptions, analyses and a discussion of the structural characteristics and the content standards of the 111 different question papers are presented. Profiles for each question paper are compared to show possible trends and patterns between years. This section also explores similarities and differences between the question papers and patterns in the styles of question papers associated with variations in the different policies which operated during this time period and which were described in Chapter 2. The second section presents descriptions and a discussion of the performance standards generated from analyses of a sample of the 2005 and 2006 candidates' scripts.[177] The third section discusses of difficulty of SC examinations from two distinct perspectives, namely, that of the macrostructure of the question papers and student performance.[178] In the fourth section, the author raises challenges in regard to the alleged equivalence within and between the 2005 and 2006 SC Biology examinations. That equivalence was assumed when students were certified in these examinations, and by the users of the results of these examinations. This chapter concludes with an exploration what the optimal length of a SC Biology examination might be and how this time period might be determined.

Space constraints and the large volume of results generated by the descriptive and comparative analyses of the question papers meant that not all the results could be presented here in the body of this text. Considerable variation was observed in the analyses of the question papers for all the examining bodies which made it difficult to choose which examining bodies to present in the main body of the text. For this reason in some instances only subsets of the results obtained from analyses of the question papers, specifically those of the CED, the WCED, the national DoE (CED-WCED-DoE) and the IEB are specifically discussed in the body of this chapter. Where necessary, the results of the analyses of the other examining bodies is used in the text to illustrate specific points. The question papers of the CED (1994, 1995), WCED (1996 to 2000) and national DoE (2001 to 2007) were selected because the sample of candidate answer scripts comes from the WCED. The CED, in part, preceded the WCED and from 2001 WCED SC students wrote the national DoE Biology SC question papers. The CED, WCED and DoE SC examinations were government regulated. The IEB was selected for inclusion in the main body of this chapter for comparative purposes because it was the one examining body that consistently set and administered its own SC examinations throughout the 14 years and because of its relative autonomy from government regulation. The results of the analyses of the other examining bodies were relegated to the appendices.[179] The relevant appendix numbers are given at the start of each section for readers interested in the complete analyses of this study. Sometimes results from the appendices are included in the discussion if appropriate examples could not be found in the analyses of the CED-WCED-DoE and IEB analyses. To aid the reader, Table 6.1 shows the figures, tables and appendices referred to in each of the sections of

177 Candidates' scripts were not available for the years prior to 2005 (see Chapter 2).

178 Due to space constraints the author does not discuss in detail the advantages and disadvantages of various test designs and format. Interested readers may consult Downing and Haladyna (2006) and Nitko and Brookhart (2007) for a discussion of test design and test formats, and Brindley (2000), Pitoniak,Young, Martiniello, King, Buteux and Ginsberg (2009) for more explicit direction in the setting of tests in a language other than that of the home language of the intended test-takers.

179 For comparative purposes the results of the CED, WCED, the national DoE and the IEB have been retained in the appendices, despite any inclusions within this chapter.

Table 6.1 Figures, tables and appendixes referred to in sections of this chapter.

Section	Section heading	Figures	Tables	Appendices
6.1	Analysis of question papers			
6.1.1	*Structure of question papers*			
6.1.1.1	Macrostructure of question papers			
a	Number, types and length of question papers		6.2 Ь 6.5	6.1 – 6.10
b	Total marks of examination		6.2 Ь 6.5	6.1 – 6.10
c	Sub-sections of question papers		6.2 Ь 6.5	6.1 – 6.10
6.1.1.2	Structure of questions and free response answers			
a	Structure of questions	6.2 – 6.4	6.6 Ь 6.13	6.11 – 6.19
b	Structure of free-response answers		6.14 Ь 6.16	6.20 – 6.28
6.1.2	*Content standards*			
6.1.2.1	Topics	6.5	6.17	6.29 – 6.37
6.1.2.2	Performance expectations		6.18 – 6.21	6.38 – 6.46
6.1.2.3	Content (topics, performance expectations)			
a	Similarity of content	6.6 – 6.7	6.22 – 6.28	
b	Breadth and depth of knowledge	6.8		6.47 – 6.55
6.1.3	*Comparison of profiles*	6.9 – 6.11		6.56
6.2	Analysis of answer scripts			
6.2.1	*Performance according to structure of questions paper*	6.12		
6.2.1.1	Choosing the correct answers or free-response answers	6.13		
6.2.1.2	Length of questions	6.14		
6.2.1.3	Use of non-text elements in questions	6.15		
6.2.1.4	Length of answers	6.16		
6.2.1.5	Non-text elements required in answers	6.17		
6.2.2	*Performance standards*			
6.2.2.1	Performance according to topic	6.18	6.29 – 6.30	
6.2.2.2	Performance according to performance expectations	6.19	6.31 Ь 6.33	
6.2.2.3	Performance according to content	6.20 – 6.25		6.57 Ь 6.58
6.2.3	*Reliability of student performance data*			
6.2.3.1	Internal reliability		6.34 – 6.41	
6.2.3.2	Factor analysis		6.42 – 6.46	
6.2.3.3	Classification of students by aggregate	6.26, 6.27	6.47 Ь 6.54	
6.3	Difficulty of examinations			
6.3.1	*Question papers*		6.55	
6.3.2	*Student performance*	6.28 – 6.36		6.59 – 6.60
6.4	Equating of examinations	6.37		
6.5	Optimal length of examinations	6.38	6.56 – 6.57	
6.6	Chapter summary	6.39		

this chapter. All results are discussed in the light of the SC Biology examinations policies (Chapter 2, Table 2.6) and of the CBS (Chapter 2, Table 2.8).

All original analyses were conducted using both the absolute marks and the absolute marks as percentages of the total marks possible in a question paper. Because the maximum number of marks can influence the difficulty of a question paper, and therefore affect comparative equivalence between question papers, the characteristics of the question papers described using absolute marks in relevant sections of this chapter. When comparison of the emphases between question papers is made, the author presents absolute marks as percentages of the total marks possible.

Note that because data on the structural aspects and the content standards were derived from data from all the question papers analyzed in this study, there was no sampling involved. Consequently no sample variance has been reported here and the issue of any statistical significance when comparing data from different examinations is generally not relevant. Any emerging contrasts between the data summaries are directly matter for educational debates.

6.1 Analysis of question papers

In their comparison of science and mathematics examinations across countries, Britton & Raizen (1996) and Britton et al. (1996a, 1996b) identified some internal features of examination question papers which characterized different examination systems around the world and which need to be considered when comparing examinations. These internal features identified included the macrostructure of examination question papers, the structure of questions asked, the structure of the answers expected, the topics covered, and performance behaviors expected from students answering the questions (Britton & Raizen, 1996; Britton et al., 1996a, 1996b). These features are analogous to those encapsulated by the structural strand and the content strand which generates the content standards (Chapter 3). This section contrasts the internal features, including the content standards, of various SC Biology examinations in several years. It forms the necessary background for understanding performance standards in eight of the examinations (Section 6.2) and provides the empirical evidence used to address the difficulty of examinations (Section 6.3), and equivalence between examinations (Section 6.4), and suggestions about the optimal length of SC Biology examinations (Section 6.5).

As previously noted, detailed analyses of the question papers indicated variability in both the structural aspects and content standards of the question papers, both within and between examining bodies and between years. As the nature and degree of the variability depends on the examination feature being considered, there are many different ways in which the analyses could be presented to make comparisons. For simplicity, the results of analyses of the structural aspects of the SC Biology examination question papers are presented first, followed by the explication of the content standards (i.e., topics and performance expectations) for these question papers. This section also includes a quantitative comparison of the content standards using a similarity index, SI. Quantitative relationships between the BOK and the DOK of each examination are then explored. Finally, profiles for each question paper are then compared within and between years using multivariate statistics.

6.1.1 Structure of question papers, structure of questions and structure of answers

Most content tests, including science tests, are to some extent language tests (AERA, APA, & NCME, 1999). The linguistic complexity of a test has been considered the most influential non-content factor on student performance because it increases the difficulty of a test, especially for test-takers who take a test in a language other than their home language (Abedi, 2006; Martiniello, 2008; 2009). This influence has made assessment of students in their non-home language "one of the thorniest difficulties in educational policy and practice" (Lee, 2005, p. 508). Unnecessary linguistic complexity is associated with construct-irrelevant threats to the intended validity of the test for all students, not just students who do not write in their home language (Abedi, 2006; Abedi & Gándara, 2006; Martiniello, 2008, 2009; Shaw, Bunch & Geaney, 2010). By careful crafting of tests, Abedi and Lord (2001), Young, Cho, Ling, Cline, Steinberg and Stone (2008) and Young, Steinberg, Cline, Stone, Martiniello, Ling, and Cho (2010) showed that it is possible to reduce the effect of English reading ability on mathematics and science tests without undermining the mathematical or scientific constructs which were being measured.

Non-linguistic components such as diagrams, tables and graphs have been shown to be important in determining the difficulty level of mathematics questions, and therefore the attained degree of validity, of mathematics tests (Martiniello, 2008, 2009) and science performance assessments (Act, 2010; Shaw, Bunch, & Geaney, 2010). Reasons given for this importance of non-linguistic elements in science are that scientific terminology is unfamiliar, that the sentence structure often used in science readings is complex, and that the mix of words, diagrams, charts, symbols and mathematics to communicate the meaning of science texts is multimodal (NRC, 2012). As non-linguistic components add to the richness of biology, they are able to "help students who have problems in reading [text]" well (Tamir, 1996, p. 71) and are therefore necessary in Biology examinations and influence how students are permitted to perform in tests.

A high proportion of South African students wrote their SC examinations in a language that was not their home language (DoE, 1998; Lolwana, 2006; OECD, 2008; Umalusi, 2004). Therefore, collecting validity evidence about linguistic complexity, such as the structures of questions and their expected answers and whether questions and answers make use of non-linguistic components, is essential in constructing a validity argument about the SC Biology examinations. In addition, general characteristics, such as test length and the number of items, influence the difficulty of a test (Abedi, 2006; Britton et al., 1996b).

The author acknowledges that evidence about the structural characteristics of SC question papers analyzed in this study does not approach the sophistication or the elegance of the studies about the linguistic complexity of tests cited above. However, such detail was beyond the scope of this thesis.[180,181,182] What this study does offer is a first exploration of the inherent non-content requirements of the SC Biology examinations and how they might relate to student performance standards (Section 6.2).

6.1.1.1 *Macrostructure of question papers*

Different formats of tests (e.g., proportion of MCQs vs open-ended questions) and test items (e.g., questions which contain non-text visual components vs those questions that are text only) have different psychometric strengths and weaknesses (Downing & Haladyna, 2006; Haladyna, 1997; Nitko & Brookhart, 2007). Prudent policy should recognize the tradeoffs between the strengths and weaknesses of different formats and promote a policy which requires a combination of different kinds of test formats to reduce the construct-irrelevant variance that comes with each specific kind of test format (Martinez, 1999). Therefore, the format of a test and its items will depend on the purpose of the test. Unfortunately, neither the CBS, nor the modified CBS which formed the basis for descriptions of macrostructure compiled for each of the SC Biology examinations question papers investigated in this study, provided reasons for

[180] The issues that non-English Language Learners (ELL) students face in an English learning environment is dependent on how syntactically or morphologically divergent the student's first-language is from English (Sousa, 2011). The linguistic research quoted here refers to mainly students from Spanish-speaking backgrounds which means that it might, or might not, apply to the South African context. I was unable to find literature pertaining specifically to different South African languages and the teaching and learning of science.

[181] The South African situation is further complicated because the SC is set and written in two of the eleven official languages, neither of which may be a candidates first language.

[182] In this study, the author made no distinction between academic language and conversational language which each influence teaching and learning in different ways (Cummins, 2003).

the specific test formats that were chosen for these examinations. It is therefore impossible to examine the rationale behind the design of the SC examinations analyzed here.[183]

What follows are descriptions of each of the variables constituting the macrostructure of the SC examination question papers which were identified in Chapter 3 as being important to understanding the standards implicit in the SC Biology examinations.

a. ***Number, types and length of question papers*** (Appendices 6.1 – 6.10)

From 1994 until 2007, the IEB and the examining bodies reporting to the DNE (1994, 1995) and to the DoE (1996 − 2000) each set and administered, per examination cycle (Chapter 2, Figure 2.3), one three-hour- long SC Biology question paper (Tables 6.2 & 6.3). From 2001 until 2007, examining bodies reporting to the DoE, including the WCED, each administered two question papers, of two hours each, which were nationally set by the DoE (Table 6.4). The IEB continued to set and administer one three-hour-long examination (Table 6.5).

For a long time it has been acknowledged that practical work, or laboratory work as it is sometimes known, is vital to the teaching and learning of science (e.g., AAAS, 1990; Black, 1996; NRC, 2012; Tamir, 1974). Despite the fact that research has shown that students perform very differently if they are asked to perform practical work than when they are asked about practical work in paper-and-pencil exercises (Black, 1996), a number of international school-leaving science examinations were found to test practical work as paper-and-pencil exercises (Britton & Raizen, 1996).[184] A danger of excluding practical work from high-stakes examinations is that teachers interpret the message that it is acceptable to teach practical work in the same way that they teach the non-practical components of science "with no serious harm" to the learning of science (Britton et al., 1996a, p. 39). As a result teachers allocate less time for practical work (Herr, 1992).

Prior to 1996, some South African examining bodies included practical work in their year mark, and from 1996 all examining bodies were required to have a year mark, allocated by the teacher, included in the final examination mark (Chapter 2, Section 2.1.3.3). During the period 1994 to 2000 a few examining bodies also included a practical test as part of their final

---

[183] The policy for the Life Sciences NSC, which replaced Biology SC examinations in 2008 uses a similar format to that used for the SC Biology examinations, also without giving reasons (DoE, 2007g).

[184] In an international study comparing school-leaving Biology examinations, Tamir (1996) reported that examinations in England, Wales, France, Germany, Japan and the USA did not have a separate laboratory practical component of their examinations, but that Israel did.

Table 6.2 Summary of marks allocated to different question types in 1994, 1995 CED and 1996-2000 WCED SC Biology HG question papers.

Year	Total mark	Section A							Section B	Section C	
		Multiple choice	Matching columns/ diagrams/ labels	Item - statement	Termin-ology/ one word	Missing words/ labels	Incorrect / correct labels	Short questions	Short questions	Data response	Essay
1994	400	22^{2} 42^{3} 21^{3}		30^{3}					225	22 (12)	38 (48)
1995	400	45^{3}	30^{3}	15^{3} 15^{3}				10	225	16 (16)	44 (44)
1996	300	40^{2}	10^{1}	30^{2}	10^{1}				165	20 (25)	23 (20)
1997	300	50^{2}	10^{1}	10^{2}	20^{2}				165	45	(45)
1998	300	40^{2}	10^{1}	20^{2}	20^{2}				165	45	(45)
1999	300	40^{2}	10^{1}	20^{2}	20^{2}				165	45	(45)
2000	300	40^{2}	10^{1}	20^{2}	20^{2}				165	45	(45)

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Numbers in brackets indicate marks for which there was a choice of questions within that section.
4. All papers were 3 hours long.

Table 6.3 Summary of marks allocated to different question types in 1994-2000 IEB SC Biology HG question papers.

Year	Total mark	Section A							Section B	Section C	
		Multiple choice	Matching columns/ diagrams/ labels	Item - statement	Termin-ology/ one word	Missing words/ labels	Incorrect / correct labels	Short questions	Short questions	Data response	Essay
1994	400	36^{3}			20^{2}			59	225		60 (60)
1995	400	36^{3}	20^{2}		12^{1}	10^{2}		37	225		60 (60)
1996	400	36^{3}						89	200 (29)		75 (75)
1997	400	36^{3}			20^{2}			69	200 (39)		75 (75)
1998	320	24^{2}				8^{1}		68	160 (20)		60 (60) (60)
1999	320	30^{2}	20^{2}					50	160 (20)		60 (60)
2000	320	18^{2}			20^{2}			62	160 (18)		60 (60)

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Numbers in brackets indicate marks for which there was a choice of questions within that section.
4. All papers were 3 hours long.

Table 6.4 Summary of marks allocated to different question types in 2001-2007 national DoE SC Biology HG and SG question papers.

Year, grade	Total mark	Section A							Section B	Section C	
		Multiple choice	Matching columns/ diagrams	Item - statement	Termin-ology/ one word	Missing words/ labels	Incorrect/ correct labels	Short questions	Short questions	Data response	Essay
HG											
2001	400	28^{2}		24^{2}	22^{1}			46	210	34	36
2002	400	34^{2}	20^{2}	6^{2}	20^{1}	8^{1}	6^{2}	26	210	34	36
2003	400	22^{2}	10^{2}	26^{2}	20^{1}			42	210	34	36
2004	400	26^{2}		26^{2}	13^{1}			55	210	34	36
2005	400	28^{2}		20^{2}	12^{1}			60	210	34	36
2006	400	30^{2}		24^{2}	16^{1}	8^{1}		42	210	34	36
2007	400	30^{2}		24^{2}	6^{1}	8^{1}		52	210	34	36
SG											
2001	300	20^{2a}	10^{2} 10^{2} 10^{2}		20^{1}			30	200		
2002	300	30^{2}	10^{2} 6^{1}	14^{2}	20^{1}			20	200		
2003	300	18^{2}	8^{2} 10^{2} 10^{2}		19^{1}			35	200		
2004	300	22^{2}	26^{2}		13^{1}		4^{2}	35	200		
2005	300	28^{2}	22^{2}		14^{1}		6^{2}	30	200		
2006	300	28^{2}	26^{2}		13^{1}			33	200		
2007	300	26^{2}	26^{2}		14^{1}			34	200		

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Data are combined for two papers, each 2 hours long.

Table 6.5 Summary of marks allocated to different question types in 2001-2007 IEB SC Biology HG and SG question papers.

Year, Grade	Total mark	Section A: Multiple choice	Section A: Matching columns/ diagrams	Section A: Item ï statement	Section A: Termin-ology/ one word	Section A: Missing words/ labels	Section A: True/ false Present/ absent	Section A: Short questions	Section B: Short questions	Section C: Data response	Section C: Essay
HG											
2001	300	14^{2}	20^{2}					66	140		60 (60)
2002	400	18^{2}			20^{2}		20^{1}	67	200		75 (75)
2003	300	20^{2}	10^{2}		10^{1}			60	140		60 (60)
2004	300	20^{2}	20^{2}					60	140		60 (60)
2005	300	28^{2}	20^{2}					52	140		60 (60)
2006	300	20^{2}	20^{2}					60	140		60 (60)
2007	300	16^{2}	20^{2}					64	140		60 (60)
SG											
2001	225	20^{2}	10^{2} 16^{2}					34	125		20 (20)
2002	300	20^{2}	10^{2}		10^{2}	10^{2}		50	175 (8)		25 (25)
2003	225	20^{2}	10^{1}		38^{2}			12	125		20 (20)
2004	225	16^{2}	10^{1}		40^{2}			14	125		20 (20)
2005	225	10^{2}	10^{1} 4^{1} 10^{2}		36^{2}			10	125		20 (20)
2006	225	14^{2}	10^{1}	4^{1}			10^{1}	42	125		20 (20)
2007	225	12^{2}	5^{1} 7^{1} 14^{2}			8^{1} 8^{1}		26	125		20 (20)

Note:
1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Numbers in brackets indicate marks for which there was a choice of questions within that section.
4. All papers were 3 hours long
5.. Both 2002 SG Section B options have identical profiles.

examination mark (Chapter 2, Section 2.1.3.3). The CBS required that students performed specific practical work and from 2001 all examining bodies were required to include practical work in a year mark, known as the CASS mark (Chapter 2, Table 2.8). External moderation of teacher and student work during 2001 to 2007 found that many teachers in government-administered schools did not require students to conduct practical work and that practical work was done as paper-and-pencil exercises (Crowe 2007a, 2007b; T. Isaac, personal communication, May 27, 2011). All question papers analyzed in this study included questions about practical work, so it is possible that many teachers addressed the teaching of practical work in the same way that they teach facts, concepts and processes.

The total number of scorable events, 11 006, which represents the units of analysis in this study is, in part, an indication of the length of question papers, as they represent sub-questions or distinct tasks (Section 5.2.1.2). The number of scorable events for each of the question papers varied considerably (Appendix 5.3). Given that scorable events represent the individual tasks required by a question paper, papers with a greater number of scorable events required that more individual tasks to be completed. However, the number of scorable events, or individual tasks, is determined by the types of question in the examination (Tamir, 1996). The IEB consistently had fewer scorable events for both HG and SG question papers than did DoE-administered question papers, when the question papers carried the same number of total marks per unit time. This structure means that individual scorable events on IEB question papers generally carried more marks than scorable events in DoE-administered question papers. The lack of correspondence between the length of question papers and the total the number of marks allocated to each scorable event was problematic in terms of the message sent to communities of educators and students about the 'value' of a mark in Biology assessments, and especially for those people charged with determining the equivalence between Biology assessments. In principle, a question paper that is less demanding in terms of its time requirements (per scorable event/question or per mark) would better enable candidates to read and process questions and to formulate their answers. This diminished time requirement would be particularly important where candidates have to demonstrate their competence in skills and processes rather than simply the recall or recognition of content. The relationship between the cognitive demand of a question paper and the total mark of a question paper was not pursued in this study because of the different values which are attached to marks for similar questions, sometimes within the same question paper. Hence the question "What is the 'value' of one mark?" is discussed in Section 6.4.

b. ***Total marks of examination*** (Appendices 6.1 - 6.10)

In South Africa, there were no policy requirements regarding the total marks of an examination until 2001, when the DoE stipulated that the two nationally set question papers should total 400 marks for HG and 300 marks for SG for students in government schools (Chapter 2, Table 2.8). Consequently, prior to 2001 the total marks of the different examination question papers varied both within and between years (Tables 6.2 & 6.3). The total marks for HG examinations in the years 1994 to 2000 in the CED-WCED and the IEB question papers ranged between 300 and 400 marks per question paper (Tables 6.2 & 6.3). In the period 2001 to 2007, when the DoE total marks for each of HG and SG question papers were constant each year (Table 6.4), IEB question papers varied between total marks of 300 or 400 for HG and total marks of 225 or 300 for SG (Table 6.5).

Ideally, for equivalence of marks between question papers to be established, a mark should carry the same value within and between question papers. If a mark has the same value within and between question papers, HG students who wrote a 400- mark question paper in three hours would have had to potentially work harder per unit time to obtain the same total percentage on their examination as HG students who wrote a 300-mark question paper in three hours. Higher Grade students who wrote 400-mark question papers in four hours had to work potentially as hard as HG students who wrote a 300-mark question paper in three hours. The statements all hold provided there is some consistent value of a mark. This again raises the question posed above about the value of a mark (See Section 6.4).

c. ***Sub-sections of question papers*** (Appendices 6.1 - 6.10)

All question papers comprised Sections A, B (HG and SG) and C (HG only), as required by the CBS (Chapter 2, Table 2.8). The proportions of marks allocated to each of Section A, B and C varied between question papers, often depending on the total marks of each question paper, but not always in accordance with the policy of the time. For example, Section C in the 1994 and 1995 HG question papers of the NEB (Appendices 6.1 & 6.2) contributed more than the 15% required by the CBS. The CED-WCED-DoE HG and the DoE SG question papers followed the CBS policy (Tables 6.2 & 6.3). The IEB SG 2001 to 2007 question papers included a Section C essay (Table 6.5) which was not required by the CBS for SG (Chapter 2, Table 2.8). This inclusion meant that the proportions of the IEB SG question papers comprised by Sections A and B differed from the CBS policy. The IEB argued that an essay question for SG allowed students the opportunity to develop an argument and enabled the examining body to differentiate better between SG candidates (James Buchanan [IEB examiner], personal communication, May 23, 2008). Unfortunately, the marks obtained by the IEB SG candidates were not archived by question, nor were scripts from students retained. This argument therefore cannot be verified. Different types of essay questions, their mark allocations, how they were marked, and the implications of these differences form part of a later discussion (Section 6.1.1.2a).

In general, Section A of both the HG and SG question paper, included a variety of different kinds, or formats, of questions: MCQs, matching columns or diagram labels, item/statement, terminology/one word, missing words/labels, identifying incorrect/correct labels, true or false and present or absent, and short questions.[185] Each type of question also contributed to Section A in varying proportions. For example, post-2000, Section A of particular DoE HG question papers included four different kinds of question (2001, 2004, 2005), whereas another (2002) included seven different kinds of questions (Tables 6.2 & 6.4). The DoE SG question papers offered between four and five different types of questions (Table 6.4). During the same time period, the IEB HG and SG question papers generally had fewer distinct types of Section A questions than did the DoE set question papers (Tables 6.3 & 6.5). In addition, the number of marks awarded to individual questions of the same format in Section A varied considerably between years, especially between 1994 and 2000. For example, in the 1994 HG examinations, MCQs in the Orange Free State question paper carried one mark each, while similar questions in the CED question paper in the same year carried either two or three marks each; the IEB question paper each MCQ was worth three marks (Appendix 6.1). Differences between the marks allocated to MCQs persisted within the same HG question papers within some years (e.g., CED 1994 [Appendix 6.1], Eastern Cape Education Department 1997 [Appendix 6.4]) and between an examining body's HG question papers in different years (e.g., CED and NEB, 1994 and 1995 [Appendices 6.1 & 6.2]). For the period 2001 to 2007, there was no distinction made within or between examining bodies with respect to marks allocated to MCQs between the HG and SG question papers, as during this period all MCQs carried two marks each. Specific examples of MCQs are discussed in Section 6.1.1.2a. Differences in mark allocation per question were observed in the other non-MCQs categories of Section A questions, and the implications of this practice is discussed further in Section 6.4.

Section A questions classified as short question type were indistinguishable from the kinds of short questions which characterized Section B. The CBS stipulated that Section A and Section B questions require short ("objective" was added in the modified CBS [DoE, 2001a]) answers. This stipulation begs the question of why the policy separates Section A and Section B? Are short questions added to Section A to simply satisfy the mark requirements of each section? What is the educational significance of this directive? Neither the CBS or the modified CBS offer an answer to either of these questions (DNE, 1984a, 1984b; DoE, 2001a).

185 These different kinds of questions are described in Chapter 5.

Generally, Section B in both HG and SG question papers comprised short questions which tested both recall and higher order intellectual skills such as application, as required by the CBS and modified CBS (Chapter 2, Table 2.8). While the CBS stipulated that Sections A and B were compulsory, they did not preclude the possibility of a choice of questions in Section B (Chapter 2, Table 2.8). The IEB offered a choice of sub-questions in Section B in HG question papers for 1996 to 2000 (Table 6.4) and SG in 2002 (Table 6.5). Given that different students have different strengths and weaknesses in terms of their competencies (Bloom, 1968), a choice of questions potentially provides students with more opportunity to demonstrate their competencies than does no choice of questions (Downing & Haladyna, 2006).

From 1994 HG Section C could have been a choice of two questions, a data-response question[186] and/or a structured essay. In 2001 the CBS was amended for the DoE-set HG Section C to have one compulsory Section C question which was half data-response question and half essay type question (Table 2.8). From 1994 to 2000, most examining bodies offered a choice of questions, in varying combinations of data-response and essay type questions, in this section of the HG examination papers. The author was unable to establish the reasons for the DET practice of replacing data-response or essay type questions in Section C with MCQs, and no choice of questions in the 1994 question paper (Appendix 6.1). In 1995, the DET retained Section C MCQs but offered students an essay as an alternative to MCQs (Appendix 6.2). The 1989, the DET question paper Section C offered a short question very like the Section B questions and provided no choice of question (Scheltema & Myburgh, 1990). Scheltema and Myburgh (1995) confirmed that in 1989 the DET started to use a format for Section C that was different to that used by the other examining bodies at the time, but it is unclear when the DET began using MCQs in Section C.[187] From 1996 to 2000, Gauteng and in 2000 KwaZulu-Natal (Appendices 6.3 - 6.7), did not offer a choice of Section C questions.

Descriptions of a selection of HG Section C questions for one year, 1996, are used here to show the kinds of variation that was observed within the examinations of one year (Scheltema & Myburgh, 1997). Any pre-2001 year would have shown similar kinds of variation. In 1996, for

186 A data-response type question is a question which uses unseen information as text, diagrams, tables or graphs about one of the topics required by the syllabus, followed by a series of sub-questions.

187 Despite approaches made to the Examinations Directorate of the national DoE, Mr Lotter (a member of the JMB), Dr Calitz (the CEO of Safcert at the time) and various people involved with the SC Biology examinations at that time, the author was unable to trace any paperwork documenting exactly when and why this change was made to the DET question papers. It could conceivably have been a decision that was made by the DET officials in acknowledgement of the inferior education that was offered by the DET at the time and an assumption that the type of MCQs offered by the DET Section C were more accessible to students than data-response or essay type questions (an assumption that is not necessarily true [Haladyna, 1997, Downing & Haladyna, 2006, Nitko & Brookhart, 2007]).

Section C a choice was offered by the Northern Province between two structured essays,[188] each addressing different topics, one relating to biochemistry, photosynthesis and human digestion and the other to nervous and chemical co-ordination in the human body. In the same year, the Eastern Cape offered a choice of a data-response type question about population dynamics and a structured essay question about human co-ordination. Mpumalanga required that, for Section C, all students answer a 30-mark data-response type question about physiological processes in plants, after which students could choose between one of two 20-mark structured essays. One essay targeted the homeostatic control of sugar in the human body and the other, the suitability of plant tissues for water movement in a plant. In the same year, the WCED offered a choice of two questions, one about photosynthesis and cellular respiration and the other about population dynamics. Each of the WCED questions had a data-response component and an essay. The IEB offered a choice of essays, each worth 75 marks. Both IEB essays required students to read a short passage and to construct an essay related to the passage. One of these essays required students to demonstrate their knowledge in two of four aspects (photosynthesis, movement of water through a plant, movement, photoperiod) of the lives of plants and the other about human thermoregulation and any one other homeostatic process.

A choice of questions, especially when the alternates are different kinds of questions, or if they address different knowledge and skills, presents challenges to equivalence. Unfortunately, no official marking memoranda or student scripts were available to examine possible effects on equivalence for the 1994 to 2000 period of this study. Additional examples of a range of essay type questions for the period 2001 to 2007 are further discussed in Section 6.1.1.2a to demonstrate not only the variability observed in essay type questions and the marks that each carried, but also the variability in how they were marked.

6.1.1.2 *Structure of questions and structure of free-response answers*

Not surprisingly, a lack of explicit direction by the CBS and the modified CBS about the structure of questions, other than what was discussed above, or with regard to how questions should be answered, resulted in considerable variation in the structure of questions and in expectations for answers. The variation in Section A, Section B and Section C question types and required answers as required by policy was discussed in Section 6.1.1.1c above. Here the analyses of the item characteristics which relate to the structure of the questions, in terms of the scorable events they comprised, is discussed. As question papers varied in terms of their total marks, the results for each item characteristic are presented here as a percentage of the total

[188] A structured essay is one in which candidates are given an explicit breakdown of the sub-sections and the marks associated with each of the sub-sections about which they are required to write.

marks possible for each question paper, in order to allow comparison of the relatives emphases of different examination question papers.

The structure of a question is inextricably linked to the answer that it is expected to elicit from a student. The sections of the discussion that follows should be understood *together* and be viewed in the light of the macrostructure of each question paper described in the preceding section. For ease of reading, this section separates the characteristics relating to the structure of questions from those related to the structure of answers.

Detailed summaries of the structure of the questions for each SC question paper can be found in Appendices 6.11 to 6.19 and the structure of the answers for each SC question paper are summarized in Appendices 6.20 to 6.28.[189] What follows are descriptions of specific features which capture some of the contrasts among the examinations analyzed in this study. Details of two kinds of questions, namely, MCQs and essay questions, are then selected to illustrate kinds of variation between question paper. These variations are then discussed in light of how they might have influenced student performance.

a. ***Structure of questions*** (Appendices 6.11 - 6.19)

How questions are formulated in tests affects the difficulty of questions and therefore affects student achievement (Downing & Haladyna, 2006; Nitko & Brookhart, 2007). It has sometimes been implied that questions which require students to choose-the-correct-answer are easier to answer because students have to recognize the correct answer rather than construct the answer as required in free-response answers (Haladyna, 1997; Nitko & Brookhart, 2007; Wood, 1977). However, not all choose-the-correct-answer questions are alike and not all free-response questions are alike (Downing & Haladyna, 2006; Haladyna, 1997; Nitko & Brookhart, 2007). Biology is a science and scientists regularly communicate by constructing and interpreting diagrams, graphs and tables (NRC, 2012) so it is to be expected that Biology examinations require students to understand such non-text in their examinations. In addition, Britton et al. (1996b) argued that the use of different visual modes within a mathematics or a science examination paper was necessary if the paper was to be fair to all students, because some students perform differently in different modes of communication.

A variety of different kinds of questions asked in the SC examinations were investigated in this study. Section 6.1.1.2c discussed the varying total marks of major question types that comprised Sections A, B and C in different examinations, and the current section describes and

[189] Values given for each paper are calculated from the sum of marks for all scorable events which comprise a particular category. A question can be comprised of one or more scorable events (Chapter 5). The numbers of scorable events identified for each paper are given in Appendix 5.3.

compares some of the more specific characteristics of question types. Those characteristics of questions included for discussion here are the length of each question (i.e., one sentence, two to three sentences, more than three sentences) and whether the question made use of non-text elements (i.e., diagrams, graphs and tables). Specific examples of MCQs and essay questions from some of the examination papers are given to illustrate a range of different types of questions and a range of possible impacts that different types might have on student performance.

The categories of sentence length used in this study were chosen because of the vast number of scorable events that had to be analyzed. Therefore, the results these measures have generated must be viewed with caution since sentences can vary in the number of words they comprise and some sentences are more complex than others. These results do however capture an element of the how much reading/processing students have to do in order to answer a question. The extent of reading is important in all student performance and cannot be captured by the number of pages,[190] the number of questions, or the number of scorable events which comprise each question paper. The ability to read comprehensively is link to college readiness (ACT, 2006).

There was a lot of variation in the length of questions comprising the examinations, but most scorable events were short (i.e., 3 sentences or less) when all the question papers from 1994 to 2007 are considered. Given that a scorable event is the unit of analysis and cannot be further broken down, one might be tempted to speculate that minimum sentence length should apply to most scorable events and therefore to a greater proportion of a question paper. However fairly large percentages of total marks (>10%) came from longer scorable events (i.e., more than 3 sentences) in some of the question papers (e.g., HG 1994 HG Transvaal Education Department – Profiles 1 and 2, House of Assembly – Profiles 1 and 2 [Appendix 6.11]; 1995 HG Natal Education Department – Profiles 1 and 2 [Appendix 6.12]; 2000 HG Northern Province – Profile 1 [Appendix 6.17]; 2002 HG IEB – Profile 2 [Appendix 6.18]; 2004 SG IEB – Profile 2, 2006 SG IEB – Profiles 1 and 2 [Appendix 6.19]).

For the 1994 to 2007 period, the percentage of longer and shorter scorable events varied for the CED-WCED-DoE and IEB set question papers (Tables 6.6 – 6.9). Patterns in these question papers were that the IEB HG questions papers had more two/three sentence long scorable events than did the CED-WCED-DoE HG question papers which had a higher proportion of one sentence scorable events. Both the 2005 DoE HG and SG question papers had a much higher proportion of two/three sentence long questions than did the other DoE-set question

190 The author has been present at South African SC examination meetings where educators have talked about the demand of a question paper in terms of the number of pages it comprises, despite the fact that often some pages contained just a few lines of text.

papers (Table 6.7). This contrast will be addressed later because of links with the cognitive demand of the question paper. Differences in the length of scorable events between the HG question papers and SG question papers (2000 to 2007) was much more pronounced for the IEB question papers where SG students had a higher proportion of shorter scorable events to answer than did the HG students (Tables 6.7 & 6.9). In 2001, DoE SG students had to respond to longer scorable events than did HG students in the same year (Table 6.7). Longer scorable events required students to read and process more information than was required by shorter scorable events. Without access to scripts of students who answered the question papers with longer scorable events, especially where a choice offered questions of different lengths (e.g., 2000 HG Northern Province, 2002 HG IEB, 2004 SG IEB), it is difficult to consider how question length influences performance. Possible effects of scorable event length on student performance will be discussed under student performance in the 2005 and 2006 examinations (Section 6.2.1.2). This set of effects links again to the question previously raised as to the value of a mark.

All question papers included questions diagrams, graphs or tables as part of their questions, but to varying degrees. The use of diagrams in question papers was more common than the use of graphs and tables. Some question papers used neither graphs or tables in questions (e.g., 1995 HG House of Representatives – Profile 1 [Appendix 6.12]), others used only diagrams and graphs (e.g., Northern Cape – Profiles 1 and 2 [Appendix 6.14]) or used only diagrams and tables (e.g., WCED – Profiles 1 and 2 [Appendix 6.15]). Diagrams and graphs only (1994 Profile 1), diagrams and tables only (1998) were used in the CED, WCED, DoE HG set question papers (Tables 6.6 - 6.7). The DoE 2005 HG question paper was characterized by the combined use of diagrams, tables and graphs in single scorable events and this feature will be addressed later because of links to the cognitive demand of the question paper. Some IEB HG examinations did not make use of any non-text in some years ─ 1995 HG (no graphs); 1999 HG, 2007 HG, 2003 SG, 2004 SG (no tables) (Tables 6.8 - 6.9).

Table 6.6 Types of questions, as a percentage of total marks, in the 1994, 1995 CED and 1996-2000 WCED SC Biology HG question papers.

Year	Profile	Total marks	One sentence	Two - three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
1994	1	400	82.3	9.8	8.0	18.0	8.5	0.0
1994	2	400	85.3	6.8	8.0	20.0	5.5	2.0
1995	1	400	78.5	20.5	1.0	17.3	13.3	7.0
1995	2	400	78.5	9.5	12.0	17.3	12.3	8.5
1996	1	300	93.3	6.7	0.0	7.7	3.7	0.7
1996	2	300	91.0	9.0	0.0	8.7	5.0	2.0
1997	1	300	83.7	13.0	3.3	20.3	4.7	10.3
1997	2	300	80.3	16.3	3.3	20.3	8.7	10.3
1998	1	300	83.0	16.3	0.7	20.3	0.0	10.7
1998	2	300	77.0	22.3	0.7	20.3	0.0	6.3
1999	1	300	87.0	12.3	0.7	18.7	8.7	9.0
1999	2	300	84.7	12.3	3.0	19.3	8.7	9.0
2000	1	300	88.7	9.7	1.7	28.3	3.3	9.0
2000	2	300	91.3	7.0	1.7	15.0	3.0	8.0

Table 6.7 Types of questions, as a percentage of total marks, in the 2001-2007 national DoE SC Biology HG and SG question papers.

Year, Grade	Total marks	One sentence	Two - three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
HG							
2001	400	98.5	0.3	1.3	28.8	6.3	7.5
2002	400	91.5	8.0	0.5	22.5	5.8	14.3
2003	400	90.3	8.5	1.3	35.3	8.8	11.0
2004	400	82.8	13.3	4.0	48.3	8.8	11.5
2005	400	55.5	40.3	4.3	32.3	9.8	18.0
2006	400	76.8	12.3	11.0	25.8	7.3	14.5
2007	400	66.5	27.3	6.3	19.8	0.5	15.5
SG							
2001	300	96.7	0.0	3.3	37.7	7.3	3.3
2002	300	93.0	5.3	1.7	48.7	6.7	3.0
2003	300	96.0	3.7	0.3	37.0	8.3	14.3
2004	300	95.7	3.3	1.0	44.0	11.0	4.0
2005	300	75.0	16.0	9.0	42.7	6.0	5.7
2006	300	84.0	11.0	5.0	47.0	8.3	3.7
2007	300	89.3	7.0	3.7	37.3	5.0	9.3

Table 6.8 Types of questions, as a percentage of total marks, in the 1994-2000 IEB SC Biology HG question papers.

Year	Profile	Total marks	One sentence	Two - three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
1994	1, 2	400	56.3	33.0	10.8	15.5	4.5	3.5
1995	1, 2	400	53.3	33.0	13.8	19.8	0.0	2.3
1996	1, 2 5, 6	400	68.0	25.8	6.3	17.3	5.8	7.8
1997	1, 2	400	64.0	32.8	3.3	25.5	6.3	1.0
1997	3, 4	400	64.0	32.8	3.3	22.0	6.3	1.0
1997	5, 6	400	64.0	32.8	3.3	28.0	4.3	1.0
1997	7, 8	400	64.0	32.8	3.3	24.5	4.3	1.0
1998	1, 2 3, 4 5, 6	320	55.0	18.4	26.6	19.7	4.4	5.9
1999	1, 2	320	73.1	24.7	2.2	19.1	1.3	0.0
1999	3, 4	320	70.6	25.3	4.1	19.1	3.1	1.9
2000	1, 2	320	49.1	43.4	7.5	31.3	3.8	3.8
2000	3, 4	320	51.6	43.4	5.0	31.3	3.8	3.8
2000	5, 6	320	47.2	45.3	7.5	31.9	3.8	3.8
2000	7, 8	320	49.7	45.3	5.0	31.9	3.8	3.8

Table 6.9 Types of questions, as a percentage of total marks, in the 2001-2007 IEB SC Biology HG and SG question papers.

Year, grade	Profile	Total marks	One sentence	Two - three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
HG								
2001	1, 2	300	57.0	29.3	13.7	19.3	8.0	4.7
2002	1	400	61.0	38.0	1.0	19.0	4.5	4.8
2002	2	400	61.0	19.3	19.8	19.0	4.5	4.8
2003	1	300	61.3	34.3	4.3	35.0	5.7	5.7
2003	2	300	61.3	34.3	4.3	27.0	5.7	5.7
2004	1, 2	300	56.3	37.7	6.0	21.7	13.0	6.7
2005	1, 2	300	64.0	32.3	3.7	25.3	11.0	8.3
2006	1, 2	300	59.0	32.7	8.3	20.3	16.7	3.0
2007	1, 2	300	61.3	29.3	9.3	23.0	9.0	0.0
SG								
2001	1	225	84.0	13.3	2.7	36.9	6.2	3.6
2001	2	225	92.9	4.4	2.7	36.9	6.2	3.6
2002	1, 2	300	87.7	8.3	4.0	45.3	9.0	2.7
2003	1, 2	225	81.8	16.4	1.8	35.6	3.1	0.0
2004	1	225	64.0	28.0	8.0	42.7	5.3	0.0
2004	2	225	64.0	19.1	16.9	42.7	5.3	0.0
2005	1	225	70.7	17.3	12.0	37.8	5.8	2.7
2005	2	225	70.7	26.2	3.1	37.8	5.8	2.7
2006	1, 2	225	68.0	14.7	17.3	11.1	6.7	13.8
2007	1, 2	225	53.8	27.6	18.7	38.7	1.3	4.9

Multiple choice questions (MCQs)

Multiple choice questions are relatively straightforward and easy to mark, increasing the reliability of test scores (Wainer & Thissen, 1993). This consequence is an important advantage of MCQs, especially in South Africa where highly skilled markers are not necessarily the norm,[191] and where large numbers of students annually write the NSC examinations, so that these questions can be objectively and efficiently marked.[192] A major disadvantage of MCQs is that they do not give examiners insight into examinees' self-expression, that is, what they can do with their knowledge unprompted by a correct answer. Such insight is necessary if reporting is required to address performance standards, other than as a single mark or symbol.[193] Multiple choice questions consistently formed a part (in varied percentages) of all the question papers investigated in this study. Much research has been conducted, especially in the USA, about how MCQs can best be constructed for valid and fair testing. Haladyna (1997) and Martinez (1999) both made the point that, contrary to popular belief amongst educators, MCQs can test students' cognitive abilities at levels beyond recall, provided that the MCQs are carefully crafted. Figures 4.14 (D) and 4.14(E) (Chapter 4) of this study confirm that MCQs in the SC Biology question papers analyzed for this study did use MCQs to test higher order skills. However, the author did note that from 2001 to 2007HG and SG question papers set by the national DoE contained very few MCQs which tested higher order cognitive skills (see Section 6.1.2.2).

Given that "effective item writers are trained, not born" (Downing, 2006, p. 11), the author has chosen to discuss the merits of different kinds of MCQ examples taken from a range of selected question papers, in the light of the current literature, with the intention of informing South African policy about the choices made about questions used in examinations. Consequently, not all the advantages and disadvantages of using MCQs in tests, nor all the criticisms levelled at the use of MCQs in tests, will be covered here. Readers interested in more details about MCQ

---

191 Knowledge of under-trained and therefore less skilled markers in the SC examinations in all subjects offered for examination has long been suggested but the author was unable to find evidence in the public domain which confirms of this view. Recently, the marks awarded to students who wrote the 2010 NSC History examination in the WCED were contested. At a meeting of affected history teachers called by the Superintendent-General of the WCED (January 20, 2011), inconsistent marking was identified as the most probable cause of students being awarded incorrect marks for the examination (R. Siebörger, personal communication, May 26, 2011). Having observed inconsistencies both within and between markers, Congdon and McQueen (2000) recommended that raters (markers) be constantly monitored, especially in large-scale testing which extends over a week or more. In South Africa, the length of time allocated to marking varies between provinces and can be longer than seven days.

192 The MCQs in Biology SC examinations are not machine-marked, they are marked by hand.

193 In Chapter 3 it is argued that performance standards are what give meaning, in terms of content standards, to a students marks. Performance standards require explanation of what students can (or cannot) do as evidenced by a particular test.

design are directed to the technical advice provided by Downing and Haladyna (2006) and Nitko and Brookhart (2007). Examples of the variety of MCQs have been selected from all examining body's question papers for three years: 1994 HG (Figure 6.2), 1996 HG (Figure 6.3) and 2006 HG and SG (Figure 6.4). Each of these years represents one of the three different eras recognized in this study (Chapter 2). Appendices 6.1 to 6.10 show that MCQs carried a variable number of marks, from 1 to 3 marks each, both within and between question papers. By investigating the selected questions, the author is attempting to determine if there is a possible explanation of the variations in mark allocation.

Multiple choice questions require students to select correct answer(s) from among the options offered but come in a variety of different formats (Nitko & Brookhart, 2007). The MCQs in this study were not of a uniform kind, nor did all MCQs carry the same mark allocation (Figures 6.2 - 6.4). Some items had four options of possible answers while others had five options, implying that the probability of weaker students guessing the answer to questions they had not mastered correctly was higher in the first configuration than in the second. Four-option MCQs are easier for random guessers. In a meta-analysis which examined 80 years of theoretical and empirical research about the optimal number of answer choices for a MCQ, Rodriguez (2005) concluded that three options was the best. Rodriquez (2005) demonstrated that different numbers of MCQ answer choices affected not only item difficulty, but item discrimination and test score reliability. Pre-1998, whether IEB MCQs had five choices of possible answers, (1994 HG [Figure 6.2 9 (C)],1995 HG) or four options (1996 HG [Figure 6.3 (C)], 1997 HG [Table 6.3]), each question counted for the same number of marks (three). This practice is not commensurate with the notion that more difficult items might carry more marks than easier options. In 1998 the IEB HG question retained MCQs with four options but decreased the mark allocation to two marks each (2006 HG [Figure 6.4 (E)]), a change which appears to have been in response to a reduction in the total marks for the question paper from 400 to 320 (1998 to 2000) and to 300 (2002 to 2007)[194] rather than because of differences in the perceived difficulty of different MCQ formats.[195]

194 In 2001, the IEB HG question paper total reverted to 400 marks and the MCQ marks remained at two marks per question.

195 In 2008, a new Life Sciences NSC curriculum replaced the Biology SC curriculum. The IEB MCQs with a choice of four possible answers were reduced to one mark each, within a total examination mark of 300 marks. In 2009, the IEB introduced differential mark allocation for MCQs, based on the comparable complexity of each question (Crowe, 2009). The DNE retained two marks for each MCQ, irrespective of the complexity of the question with a total of 300 marks for the examination (Crowe, 2009).

A. Cape Education Department

1.1.1 Which of the following combinations can complete the following statement correctly?

The importance of water in nutrition is that it serves as a:

(i) catalyst in condensation reactions
(ii) medium for chemical reactions
(iii) transport medium for nutrients
(iv) reserve energy

A (i) and (ii)
B (i) and (iii)
C (ii) and (iii)
D (i) and (iv) (3)

1.1.2 Which of the following combinations represents plant macro-nutrients?

A Iron, calcium and phosphorus
B Sodium, calcium and nitrogen
C Iodine, magnesium and calcium
D Calcium, nitrogen and phosphorus (3)

1.2.3 Which TWO of the following substances are transported by the lymph vessels to the blood?

A Uric acid
B Fats
C Proteins
D Hormones (2)

1.4.3 Which of the substances listed below cannot be digested further?

A Sucrose
B Amino acids
C Glucose
D Emulsified fats (2)

B. Orange Free State Education Department

1.3.7 The THREE hormones directly involved in the regulation of the blood sugar concentration in the bloodstream of humans, are é

A insulin, adrenalin, glucagon
B adrenalin, thyroxin, insulin
C glucagon, thyroxin, secretin
D secretin, insulin, thyroxin (1)

C. Independent Examinations Board

1.3.10 Why do proteins have a buffering effect on cells? They é

A are non-polar.
B form colloidal solutions.
C are major components of plasma membranes.
D contain basic amino acids.
E are amphoteric (3)

1.3.11 Which of the lines in the table show statements which are true of polysaccharides in living organisms?

	provide energy	*form storage compounds*	*form structural compounds*	
A	no	no	yes	
B	no	no	no	
C	yes	no	no	
D	yes	yes	no	
E	yes	yes	yes	(3)

Figure 6.2 Examples of HG multiple-choice questions and their associated marks within and between different examining bodies in 1994.

A. Western Cape Education Department

1.3.4 Take note of the following four statements:

(i) Loss of water occurs.
(ii) Causes cooling.
(iii) Excretion takes place.
(iv) Increases in hot weather.

Which of the above concern both the <u>sweating process</u> in animals and the <u>transpiration</u> <u>process</u> in plants?

A (i), (ii), (iii) and (iv)
B (ii), (ii) and (iii) only
C (i), (ii) and (iv) only
D (iii) and (iv) only (2)

1.3.6 Auxins are é

A catalysts for the process of respiration.
B vitamins needed by plants.
C growth substances causing tropisms.
D enzymes found in the gastric juice of herbivores. (2)

B. Northern Cape Education Department

1.3.2 Aerobic respiration is more advantageous to a large organism than anaerobic respiration because aerobic respiration

(i) does not require molecular oxygen and hydrogen
(ii) releases more energy from an equal amount of nutrients
(iii) produces oxygen as a waste product
(iv) does not require sunlight

A (i) and (ii)
B (ii) and (iii)
C (ii) only
D (i) and (iv) (2)

6.11 During cellular respiration oxygen is necessary to é

A break pyruvic acid down to PGA.
B combine with hydrogen during oxidative phosphorylation.
C to break glucose down to PGA.
D form coenzymes. (3)

C. Independent Examinations Board

1.3.2 The hormone secretin is produced in response to é

A acid chyme entering the duodenum
B pancreatic juice entering the ileum
C peristalsis in the ileum
D bile entering the duodenum (3)

1.3.12 An experiment was carried out on rats to find the effects of vitamin C and of protein on the growth of a group of test animals. There were four combinations of treatments. The data were set out in the table shown, with letters replacing numbers.

	Mean mass of N animals	
Treatment	Low vitamin	High vitamin
Low protein	*a*	*b*
High protein	*c*	*d*

Which of these combinations of values would provide the most information about the effect of vitamin levels on the growth of the animals?

A $\frac{a + b + c + d}{4}$

B $\frac{(a + c) - (b + d)}{2}$

C $\frac{(a + b) - (c + d)}{2}$

D $\frac{(a + d) - (b + c)}{2}$ (3)

Figure 6.3 Examples of HG multiple-choice questions and their associated marks within and between different examining bodies in 1996.

A. National Department of Education 2006 HG Paper 1

1.1.1 A function of the liver is to é

A produce bile.
B secrete proteins.
C reabsorb water.
D absorb digested food. (2)

1.1.5 Which of the following functions are performed by the substance labelled P [part of a leaf] ?

(i) Prevents cells from drying out
(ii) Protects against mechanical injury
(iii) Allows gases to diffuse in a dissolved state
(iv) Facilitates breathing movements

A (i) and (ii)
B (i) and (iii)
C (i), (ii) and (iv)
D (i), (ii), (iii) and (iv) (2)

B. National Department of Education 2006 SG Paper 1

1.1.6 Germinating seeds have a high rate of respiration because they é

A are lacking leaves.
B require energy for rapid growth.
C photosynthesise slowly.
D consist of a few cells only. (2)

C. National Department of Education 2006 SG Paper 2

1.1.5 Plant growth substances é

(i) only promote growth.
(ii) only inhibit growth.
(iii) stimulate and inhibit growth.
(iv) are only produced in the leaves.

A Only (i) is correct
B Only (iii) is correct)
C (i) and (iv) are correct
D (ii) and (iv) are correct (2)

1.1.6 Nitrogenous wastes are removed from the body by the é

A kidneys and skin only.
B kidneys, skins and the liver.
C kidneys and the liver only.
D kidneys only. (2)

\

D. Independent Examinations Board 2006 HG

8. The Two-leaf Hakea is a plant found in South-Western Australia, where spring is relatively cool and wet but summer is very hot and dry. The table shows the average values of a range of measurements taken from leaves.

Characteristic	A	B
Length (mm)	33	55
Maximum width (mm)	10	0.8
Surface area (mm^2)	288	145
Volume (mm^3)	64	63

The surface area to volume ratios of leaf A and leaf B are:

A 4.5 : 1.0 and 2.3 : 1.0
B 1.0 : 4.6 and 1.0 : 2.3
C 0.2 : 1.0 and 0.4 : 1.0
D 1.0 : 0.2 and 1.0 : 0.4 (2)

Figure 6.4 continued on next page

10. Which of these statements is **incorrect**?

A The higher the temperature the higher the transpiration rate.
B The lower the relative humidity the higher the transpiration rate.
C Relative humidity and temperature have opposite effects on the rate of transpiration.
D Relative humidity and temperature are both directly proportional in their effect on the rate of transpiration. (2)

E. Independent Examinations Board 2006 SG

1g.5 Which of the processes helps to break down food, as well as move food along the gut?

A Ingestion
B Peristalsis
C Digestion
D Absorption (2)

1g.6 Which row in the table correctly shows the final products of digestion of carbohydrate, fat and protein?

	Carbohydrate	Fat	Protein
A	Glucose, sucrose	Fats, fatty acids, glycerol	Amino acids
B	Glucose, fructose, sucrose	Fatty acids, glycerol	Amino acids
C	Glucose, fructose	Fatty acids, glycerol	Amino acids
D	Glucose, fructose, sucrose	Fats, fatty acids, glycerol	Amino acids

(2)

Figure 6.4 Examples of HG and SG multiple-choice questions and their associated marks within and between different examining bodies in 2006.

Different kinds of MCQs required varying extents of mental processing with respect to the science being tested and to how each question was asked. Some questions, for example, Question 1.1.1 (Figure 6.2 [A]), Question 1.3.2 (Figure 6.3 [B]), Question 1.1.5 (Figure 6.4 [A]), and Question 1.1.5 (Figure 6.4 [C]) required students to consider combinations of responses. Some seemingly more complex[196] questions carried three marks each and others two marks each – mark allocations that bear no obvious relationship to the total number of marks for each question paper or to other MCQs within the same question papers. For example, the CED 1994 HG question paper included more complex MCQs (Question 1.1.1, Figure 6.2 [A]) and other less complex MCQs in the *same* question paper worth three marks each (Question 1.1.2, Figure 6.2 [A]), and other less complex MCQs worth two marks each (Question 1.2.3 and Question 1.4.3, Figure 6.2 [A]). Within this question paper students had to select two correct options on one MCQ (Question 1.2.3, Figure 6.2 [A]) and in another MCQ one of four possible options (Question 1.4.3, Figure 6.2 [A]), both for two marks each. In the same year, 1994, Orange Free State MCQs carried only one mark each (Figure 6.2 [B]) with a total of 300 marks for the question paper, while the IEB MCQs carried three marks for MCQs which required students to recall similar amounts of information (Questions 1.3.10 and 1.3.11, Figure 6.2 [C]), within a total of 400 marks for the paper. Similar differences in the marks allocated to MCQs

[196] Here 'complex' here refers to the complexity of the science of the question and/or the complexity of how the question was asked and/ or how the possible answers were presented.

within the same HG question papers (Figure 6.3 [A, B, C]; Figure 6.4 [A, C, D]) or the same SG question papers (Figure 4.3 [C]) were evident in other years, also without any observable reason for the allocating of marks. There also appears to be no obvious difference between how marks were allocated to MCQs in HG and SG question papers. In the SG question papers, less and more complex MCQs were all awarded two marks each (Figure 6.4 [B, C, E]). This again raises the question about the value of a mark. See Section 6.4.

Essay questions

Britton et al. (1996a) defined an essay as a single question with no sub-questions, allotted 20 minutes or more, that required more writing or problem solving than other free-response items in the examinations. Nitko and Brookhart (2007) extended this definition by recognizing two types of essays: first, restricted response items which limit or guide the content and format of students' answers; second, extended response items in which students have the freedom to express their own ideas and inter-relationships between ideas and to determine their own organization of their answer. The CBS and the modified CBS examination policy required the presence of a "structured essay" or a "mini-essay" in Section C of the HG examination (Chapter 2, Table 2.8). While the CBS did not explain what was meant by a structured essay, essays of both kinds described by Nitko and Brookhart (2007) were observed in the HG question papers.

A selection of essays taken from examinations during 2001 to 2007, a period when the IEB still operated under the CBS policy of a structured essay, and the national DoE under the modified CBS policy which required a mini-essay, is used to demonstrate the variety of essays that were set and the different ways that essays were marked during this time period. All the essays set by the national DoE during this time-period (Table 6.10) use the words "describe", "explain" and "discuss" which give students very little opportunity to express a point of view. Most of the marks allocated by the DoE in memoranda for essays were for factual information (Table 6.11 [A]) which means students who simply regurgitated *any* 15 relevant facts *in any order* would be awarded at least 83% (15 out of 18 marks) for their essays. Analyses of students scripts revealed that many did.[197] By comparison, several of the IEB HG essay topics gave students the option to express their own views (Table 6.12), and even when descriptions were required

[197] The use of the phrase 'any x marks' (generally one mark per fact) in memorandum answers was not restricted to essay questions. When there was more information possible than was specifically required in free-response answers 'any x marks' was included in the memorandum answers. A student was awarded marks for any factual/conceptual/procedural information presented, irrespective of whether or not the factual/conceptual/procedural information was connected in a meaningful way in his/her answer. A danger of this practice is that if teachers teach to the SC Biology examinations which do not reward evidence of an understanding of the factual/conceptual/procedural information in answers, they will teach topics as sets of

the mark allocation was such that a student with all the relevant facts could get at most 40% (24 out of 60 marks) for that question, if the logical order of the facts was flawed (Table 6.11 [B]). The IEB SG essay topics (Table 6.12), while wordy, were similar to the DoE HG essay topics in that they required students to "describe", "explain" and "discuss" but an IEB SG student who had all the relevant facts, but little or no coherence, could at most get 65% of the marks (13 out of 20 marks) (Table 6.11 [C]). Clearly the IEB placed more emphasis on students making relevant and coherent connections between the factual information recalled. The memoranda for questions set in the style of the DoE HG essays could have been re-structured, at least to force a balance between the facts required, by making some combination of facts non-negotiable and others optional. The two different ways to mark essays as used by the national DoE and the IEB illustrate inherently different values of a mark awarded in an essay type question.

Given the nature of the DoE essay-type questions in Section B, and how these questions were marked, one may question of Section B and Section C were required to be separate sections in the CBS (Chapter 2, Table 2.8). One might question if this practice intended to simply satisfy the mark requirements of each section, or the educational significance of this directive in the CBS policy.

b. ***Structure of free-response answers*** (Appendices 6.20 - 6.28)

In an analysis of the kinds of questions which comprised the high school leaving examinations of a group of international countries, Tamir (1996) found that free-response questions counted for at least one-third of the total examination score. In South Africa, free-response questions counted for much more.[198] Free-response answers were the most frequently used item type in every examination, both HG (60.0% to 93.9% of a question paper) and SG (60.0% to 86.2% of a question paper), except for the 1994 and 1995 (Profile1) DET HG examinations, where choosing the correct answer was more emphasized (Appendices 6.20 & 6.21). Not all free-response items are alike (Downing & Haladyna, 2006; Nitko & Brookhart 2007). Depending on the nature of the question and the students' knowledge, the answer will vary in length and, in Biology, could also involve the student drawing diagrams or graphs or composing tables. The HG question papers which gave students the most opportunity to construct their own answers (more than 90% of question paper) were the Natal Education Department and the House of

disconnected facts, lacking any logical organization into the concepts and into larger ideas which characterize the nature of modern biological knowledge.

198 Britton et al. (1996a) made a distinction between free-response items that can be objectively marked and those that cannot. In the post-2001 SC Biology examinations, all free-response items, including essays, were objectively scored (Appendix 2.1, Item 20). In cases where a free-response answer was dependent on a previous answer, provision was made in the memoranda for alternate answers, provided the integrity of the science was maintained.

Table 6.10 Essay topics for the 2001-2007 national DoE Biology HG question papers.

Year	Paper	Topic
2001	Paper 1	Describe how the human body regulates the blood sugar level after a breakfast of maize meal porridge.
	Paper 2	Describe the movement of a molecule of water from the xylem of the leaves through the pore on the surface of the leaf as water vapour.
2002	Paper 1	Discuss the breathing pattern of an athlete at rest, how this changes during a race and immediately after the race (the recovery period) as illustrated in the sketches below [three stick people ï at rest, during the race, recovery period].
	Paper 2	Explain how a person who is running and is suddenly frightened by the sound of a hissing snake, responds.
2003	Paper 1	The normal blood glucose level in the human body is maintained at approximately 0[.]7 g per cm^3. Describe the relationship between the pancreas and the liver in maintaining the glucose of the blood at this level.
	Paper 2	Explain homeostatic control of the body temperature of a mammal under environmental conditions which cause the dilation of the blood vessels in the surface of the skin.
2004	Paper 1	Describe the process of photosynthesis from the time light is absorbed until carbohydrates are formed.
	Paper 2	Describe the role of adrenalin in preparing the person in Diagram 1 [person responding to a snake] to respond to the situation he finds himself in.
2005	Paper 1	The higher the concentration of CO_2 created in the cells and the tissue fluid as a result of more exercise, causes a drop in the pH so that it becomes slightly acidic. This situation is not favourable for the functioning of the organism as a whole. Describe the mechanism by which the breathing system will bring about a homeostatic balance to correct this situation.
	Paper 2	A person hears a car moving out of control and crashing into another vehicle. Describe the events that will lead to the person hearing the crash.
2006	Paper 1	Discuss the mechanism of breathing as it occurs in the human body.
	Paper 2	The root hair is structurally suited for its function. Explain this statement in the light of the process by which the root hair absorbs water from the soil. Describe how the absorbed water is then transported to the xylem of the root.
2007	Paper 1	In plant cells a certain chemical process releases oxygen while another uses up this gas. Refer to these processes respectively and describe ONLY the phase during which plant cells release oxygen and the phase during which plant cells use up oxygen.
	Paper 2	During Metabolism waste products are formed in the human body. It is important that these waste products be excreted from the body since they are toxic. Explain the role of the kidney in excretion by describing the role of the nephron in urine formation.

Note: 1. All essays carried 18 marks each.
2. There were no SG essays.

Table 6.11 The rating scales used for the marking of essay-type questions in the 2001-2007 SC Biology examinations.

A. National DoE HG 2005

Mark	Assessment Criteria
15	Factual content: 15 facts any order
3	Synthesis: Marks allocated as follows: 0 Not attempted / if flow charts given 1 Significant gaps in the logic and flow of the answer 2 Minor gaps in the logic and flow of the answer 3 Well-structured ï demonstrates insight and understanding of the question

Note: The wording describing the assessment criteria differed slightly between the years 2001 to 2007, but the mark allocations remained relatively the same.

B. IEB HG 2005

Marks	Assessment Criteria
24	Accuracy, relevance and completeness of content
28	The skill with which you use this factual content to answer the question
8	The way in which you organize your work, and the neatness and legibility of your presentation

Note:
1. The wording describing the assessment criteria differed slightly between the years 2001 to 2007, but the mark allocations remained relatively the same.
2. In 2002 the essay mark counted 75 marks , proportionally allocated according to the criteria used in 2001, and 2003 to 2007.

C. IEB SG 2005

Marks	Assessment Criteria
13	Factual information on the topic
4	Relevance, accuracy and focus
3	Neatness and general layout of the article

Note:
1. The wording describing the assessment criteria differed slightly between the years 2001 to 2007, but the mark allocations remained relatively the same.
2. In 2002 the essay mark counted 25 marks, proportionally allocated according to the criteria used in 2001, and 2003 to 2007.

Table 6.12 Essay topics for the 2001-2007 IEB SC Biology HG and SG question papers.

Year	Grade	Topic
2001	HG	**Option 1:** ñA large surface area to volume ratio is an essential feature of organelles in cells, of whole organisms as well as of organs in organisms.ò By referring to at least TWO different metabolic processes state whether you agree or disagree with this statement. **Option 2:** Some processes which are needed for an organism to stay alive could be harmful to that organism if they were not controlled in some way. Discuss the strategies used by plants and animals to reduce the damage which could be caused by these processes.
	SG	**Option 1:** The heart is the organ in the body responsible for ensuring that blood is effectively pumped to the lungs to get oxygen as well as to the rest of the body to make sure that all cells in the body get an adequate supply of this oxygen. Explain how blood which is transported to the heart, by both the venae cavae, is pumped **through the heart to the lungs and back to the heart again until it enters the aorta**. Your explanation should include information on how the heart is suited to this task as well as the route the blood takes. **Option 2:** Describe the different features of a leaf which enable it to carry out its functions of photosynthesis effectively.
2002	HG	**Option 1:** ïPlants are adapted to make use of the physical and chemical properties of water to provide for the needs of the whole organism.ò Use examples of THREE processes in plants to discuss whether you agree or disagree with this statement. **Option 2:** There are many involuntary processes needed in a mammalian body to make sure that life carries on. Some of these are irregular and short-lived, but others are regular and happen all the time. Both types need to be co-ordinated within an organism. Discuss the role of the **nervous and endocrine systems** in the **co-ordination** of these involuntary processes.
	SG	**Option 1**: Water is an essential substance for the survival of flowering plants and mammals. Discuss FOUR ways in which **either** flowering plants **or** mammals need water for their survival. **Option 2:** During strenuous exercise, the human body temperature rises, which is very dangerous. Explain how the body responds to ensure that the body temperature is kept within safe limits.

Table 6.12 continued

Year	Grade	Topic
2003	HG	**Option 1:** The principle of a biological control system is show in the figure [diagram of feedback mechanism]. Many human systems are controlled in this way. Write an essay in which you compare and contrast the control of body temperature with that of metabolic rate, using the information in the diagram. **Option 2:** ñFlowering plants are highly specialized to obtain the raw materials they need for photosynthesis.ò Choose any TWO of these raw materials and explain how the plant is suited to getting them.
	SG	**Option 1:** An important principle in Biology is that structure is related to function. Describe the ways in which the structure of the lungs and associated air tubes are well suited for their functions of carrying clean air to the lungs and of gas exchange. **Option 2:** Reflex actions are important to protect the body from harm. Describe an example of a reflex action that protects the body and show how the structure of the reflex arc in the body causes it to happen effectively.
2004	HG	**Option 1:** ñIn order to stay alive, living organisms need to exchange gases between their environment and their body cells.ò Write an essay in which you describe the ways in which the human body and flowering plants obtain the gases they need and are suited to doing this effectively. **Option 2:** ñWithin the human body, the composition of the internal environment is constantly and closely regulated so that homeostasis can occur.ò Describe the role of the nervous and endocrine systems in this regulation. Give at least THREE different examples of how the composition of the internal environment is kept within narrow limits.
	SG	**Option 1:** A new pupil has arrived in your class. He does not understand how the digestive system works to process the food we eat and why it is important for good health. Write a letter to him in which you explain the parts of the human gut and how each helps in the digestion of carbohydrates. **Option 2:** This picture [photograph] shows Hestrie Cloete proudly carrying the South African flag as she does her ñlap of honourò after winning the Womenõs World High Jump Championship in 2003. Her winning jump meant that her muscles had to expend a huge amount of energy to lift 68 kg (her mass) up an over 2 metres [meters]. Someone told her that this energy came all the way from the sun! Write a letter to Hestrie in which you explain how energy from the sun is trapped by plants and is then released during respiration in her muscles to be used to lift her off the ground.

Table 6.12 continued

Year	Grade	Topic
2005	HG	**Option 1:** ñMetabolic processes in living organisms produce many waste products, some of which are toxic if allowed to accumulate.ò Write an essay in which you refer to TWO toxic waste products, and discuss whether the body is well suited to remove them. **Option 2:** ñAlthough green plants are able to manufacture their own food they are not totally self-sufficient, but rely on the external environment to provide them with a variety of substances.ò Describe in what ways plants need any THREE of these substances, and discuss how well adapted you think plants are to obtain them from the environment.
	SG	**Option 1:** Hypothermia is the name given to the medical condition when you are unable to keep up your normal body temperature so that the body becomes cold. If it is not treated quickly a hypothermic person dies. Although it is unusual in South Africa, there are up to 1 000 cases in Britain each year and many more in Northern Europe, Asia and America. Imagine a fisherman has died of hypothermia after falling into the sea in the cold waters of the South Atlantic off Cape Town. Your local newspaper asks you to write an article about it. Write a newspaper to explain to readers how the body prevents its temperature from falling too low and why it is dangerous if hypothermia occurs. **Option 2:** Smoking tobacco products can cause cancer of the lungs and many others serious problems such as frequent bronchitis, shortness of breath and heart failure. Your local newspaper has asked you to write an article about the problems caused by smoking. Write an article for the newspaper to explain to readers about the structure of healthy lungs and the breathing system and how smoking is harmful.
2006	HG	**Option 1:** ñIt is essential for all living human body cells to receive a constant supply of glucose.ò Write an essay in which you discuss the need for the control of glucose levels in the blood. Then, describe how absorption in the human alimentary canal and subsequent hormonal influences enables this control to occur. **Option 2:** ñPopulation growth rate is a product of the interaction of parameters influencing a particular population.ò Write an essay in which you analy[z]e this interaction as it would relate to a population of migrating buck in open grasslands. Explain how these dynamics allow the buck population to be sustained at its carrying capacity.

Table 6.12 continued

Year	Grade	Topic
		Option 1:
2006	SG	Young cattle on a farm are dying mysteriously. They grow well for a year after they have been born. After that, although they eat normally, they do not get larger but instead start to lose weight and die. The local vet discovered that the gut of the young cattle had not started to develop properly. The small intestines were short and smooth on the inside, and there were no villi or folds. The editor of the local newspaper found out that the structure of the small intestine of these young cattle is identical to that of humans. The editor also knows that you have taken Biology in Grade 12 and asks for your help. He wants you to write a newspaper report explaining why these young cattle have died. Write a report headed ñThe mysterious death of young cattleò.
		Option 2:
		The editor of the local newspaper loves sports. In summer, he likes doing sport. In winter, he likes watching sport on TV (television). However, the editor is puzzled by something. He thought that each day he should produce about a litre of pale urine, i.e., the same amount and colour of urine. What he finds is that in summer he produces just a little urine that is dark in colour, and in winter he produces a lot of urine that is pale in colour. The editor wants you to write a newspaper report explaining why a personòs urine colour and amount change between summer and winter. He also wants you to include information on the effect of what a person drinks and does on urine production. Write a report headed ñDaily urine, why it changes in winter and summerò.
2007	HG	**Option 1:**
		óWater, water everywhere but not a drop to drinkô Write an essay in which you describe how flowering plants cut down on water loss in **dry** conditions and how the human kidney assists in preventing excess water loss when the body is **dehydrated**.
		Option 2:
		óEndurance exercise, such as long distance running, may cause dehydration and a shortage of oxygen in the human bodyô Write an essay in which you discuss how the kidney and hormones function to reduce dehydration while the breathing system adapts its activity to attempt to provide more oxygen.
	Standard Grade	**Option 1:**
		Athletes will be playing many matches in the Soccer World Cup. Write a newspaper article in which you explain the biological importance of carbohydrates, lipids (fats) and proteins in the bodies of these athletes. Choose a suitable title for your article.
		Option 2:
		Write an article for the gardening column of a newspaper about how important water is in keeping plants alive. Choose a suitable title for your article.

Note: 1. In 2001 and 2003 to 2007, all IEB HG essays carried 60 marks each and SG essays carried 20 marks each.
2. In 2002, IEB HG essays carried 75 marks each and SG essays carried 25 marks each.

Assembly (1994 and 1995), the IEB (1994, 1997, 1998, 2000) and KwaZulu-Natal Education Department (1999, 2000). The proportion of free-response items was lower for the CED and WCED HG question papers (1994 to 2000) (Table 6.13) than for the DoE HG set questions papers from 2001 to 2007 (Table 6.14). Generally the IEB HG question papers required more free-response answers than did the national DoE HG question papers (Tables 6.14 and 6.15).

The emphases on free-response questions in IEB HG examinations between years varied by as much as 13% (between 2002 and 2005) (Table 6.15). Similar relative emphases on free-response answers in HG and SG examinations were observed for the national DoE (Table 6.14) and the IEB (Table 6.16) but the IEB showed more varied emphases on free-response answers within and between HG and SG question papers (Table 6.16). For example, the IEB 2002 HG and 2004 HG examinations had an additional 9.7% and 1.4% more free-response questions respectively than did their SG examinations in the same years.

Table 6.13 Types of free response answers, as a percentage of total marks, in the 1994, 1995 CED and 1996-2000 WCED SC Biology HG question papers.

Year	Pro-file	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
1994	1	400	28.8	71.3	14.0	28.8	25.0	3.5	0.0	0.0
1994	2	400	28.8	71.3	13.0	22.3	31.5	3.5	1.0	0.0
1995	1	400	30.8	69.3	25.8	27.0	15.0	2.0	0.0	0.0
1995	2	400	30.3	69.8	26.8	26.0	15.0	2.0	0.0	0.0
1996	1	300	26.7	73.3	15.0	15.0	43.3	0.0	0.0	0.0
1996	2	300	26.7	73.3	16.3	17.3	39.7	0.0	0.0	0.0
1997	1	300	25.7	74.3	15.0	32.0	25.7	0.0	1.7	0.0
1997	2	300	25.7	74.3	16.3	37.0	19.3	0.0	1.7	0.0
1998	1	300	26.3	73.7	21.0	24.7	23.3	2.0	2.7	0.0
1998	2	300	26.3	73.7	19.7	20.3	31.7	2.0	0.0	0.0
1999	1	300	25.0	75.0	29.3	23.7	22.0	0.0	0.0	0.0
1999	2	300	25.0	75.0	33.7	28.3	13.0	0.0	0.0	0.0
2000	1	300	27.0	73.0	23.7	29.7	17.7	2.0	0.0	0.0
2000	2	300	27.0	73.0	18.3	25.3	27.3	2.0	0.0	0.0

Table 6.14 Types of free response answers, as a percentage of total marks, in the 2001-2007 national DoE SC Biology HG and SG question papers.

Year, grade	Pro -file	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
HG										
2001	1	400	13.0	87.0	26.0	37.5	22.3	0.0	1.3	0.0
2002	1	400	22.0	78.0	24.3	23.3	19.3	3.8	4.5	3.0
2003	1	400	18.3	81.8	22.0	38.0	17.0	1.8	2.0	1.0
2004	1	400	14.0	86.0	19.8	39.0	18.3	2.5	5.5	1.0
2005	1	400	16.3	83.8	13.3	44.0	19.3	3.5	2.8	1.0
2006	1	400	19.8	80.3	18.3	31.3	22.3	0.8	6.0	1.8
2007	1	400	16.8	83.3	19.5	36.0	20.3	1.5	6.0	0.0
SG										
2001	1	300	16.7	83.3	39.7	26.0	11.3	5.0	0.0	1.3
2002	1	300	21.3	78.7	42.3	26.3	8.0	2.0	0.0	0.0
2003	1	300	22.0	78.0	40.3	28.0	5.3	4.3	0.0	0.0
2004	1	300	18.3	81.7	37.3	28.3	10.0	4.7	0.0	1.3
2005	1	300	17.0	83.0	30.3	42.3	8.0	2.3	0.0	0.0
2006	1	300	23.3	76.7	26.3	40.3	4.3	3.3	0.0	2.3
2007	1	300	22.3	77.7	28.0	33.7	6.7	6.0	0.0	3.3

Table 6.15 Types of free response answers, as a percentage of total marks, in the 1994-2000 IEB SC Biology HG question papers.

Year	Pro-file	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
1994	1, 2	400	9.3	90.8	14.0	21.5	49.8	1.0	3.0	1.5
1995	1, 2	400	14.0	86.0	11.5	10.3	47.0	12.5	3.0	1.8
1996	1, 2,	400	10.5	89.5	17.8	17.0	51.5	0.8	2.5	0.0
1996	5, 6	400	10.5	89.5	17.8	14.8	53.8	0.8	2.5	0.0
1997	1, 2	400	9.0	91.0	27.3	25.8	35.5	2.5	0.0	0.0
1997	3, 4	400	9.0	91.0	21.8	28.0	35.5	5.8	0.0	0.0
1997	5, 6	400	9.0	91.0	26.3	25.3	37.0	2.5	0.0	0.0
1997	7, 8	400	9.0	91.0	20.8	27.5	37.0	5.8	0.0	0.0
1998	1, 2, 3, 4, 5, 6	320	9.7	90.3	10.0	26.9	51.3	2.2	0.0	0.0
1999	1, 2, 3, 4	320	15.6	84.4	19.1	17.5	42.8	1.3	1.3	2.5
2000	1, 2, 3, 4	320	5.6	94.4	24.7	17.5	47.2	5.0	0.0	0.0
2000	5, 6, 7, 8	320	5.6	94.4	25.9	19.4	44.1	5.0	0.0	0.0

Table 6.16 Types of free response answers, as a percentage of total marks, in the 2001-2007 IEB SC Biology HG and SG question papers.

Year, grade	Pro-file	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
HG										
2001	1,2	300	13.3	86.7	22.3	17.3	39.0	7.3	0.7	0.0
2002	1, 2	400	5.0	95.0	28.0	13.3	50.8	3.0	0.0	0.0
2003	1, 2	300	6.7	93.3	22.3	26.0	40.3	2.7	0.0	2.0
2004	1, 2	300	13.3	86.7	23.0	27.7	33.3	2.7	0.0	0.0
2005	1, 2	300	17.7	82.3	18.0	22.7	37.0	3.3	1.3	0.0
2006	1, 2	300	13.3	86.7	7.3	31.3	39.7	6.3	0.0	2.0
2007	1, 2	300	12.3	87.7	12.7	32.7	35.3	7.0	0.0	0.0
SG										
2001	1,2	225	21.3	78.7	26.2	28.4	22.2	0.9	0.9	0.0
2002	1,2	300	14.7	85.3	22.0	26.3	37.0	0.0	0.0	0.0
2003	1, 2	225	13.8	86.2	45.8	16.9	21.8	1.8	0.0	0.0
2004	1, 2	225	14.7	85.3	46.2	18.2	19.1	0.0	1.8	0.0
2005	1, 2	225	21.8	78.2	44.4	11.6	20.9	1.3	0.0	0.0
2006	1, 2	225	22.2	77.8	18.2	42.7	16.9	0.0	0.0	0.0
2007	1, 2	225	40.0	60.0	24.4	19.6	14.7	1.3	0.0	0.0

Some question papers did not require students to communicate any of their answers as diagrams, graphs or tables (e.g., WCED HG 1996 and 1999 [Table 6.13]); IEB SG 2002 [Table 6.16]). This omission is surprising given the importance placed on non-text visual communication in the sciences (NRC, 2012). “Mathematics is the “language of science,” and facility with quantitative problem solving and interpretation [and production] of graphs bestows an advantage on students” (Schwartz et al., 2009, p. 813). The marks allocated for questions requiring students to draw a graph, independent of requiring the student to show evidence of understanding the content of the graph, increased from a few marks (e.g., four marks in CED 1994 HG) pre-2001 to as many as 13 marks per graph post-2000 (e.g., DoE 2007 HG). In the DoE HG 2006 and HG 2007 question papers the drawing of graphs contributed 6% of the total mark of the examinations (Table 6.14). Drawing graphs is considered in this study as Perform-Routine-Procedures (i.e., classified as a LOCS) (Figure 4.5) but this activity was classified by the DoE as application questions which they interpret as a HOCS giving rise, in part, to differences in the weightings of LOCS and HOCS observed in Figure 4.11 (Chapter 4). Given that the drawing of graphs is a skill that was required by Grade 9 Natural Science (Luckay, 2010) and Mathematics (S. Jaffer, personal communication, July 27, 2011) students, and that the drawing of graphs requires no demonstration of knowledge specific to Biology, should there be such emphasis on simply drawing graphs independent of content in the SC examinations? In what sense can marks for graphs be interchangeable with marks for text or tabular equivalents? The level of mathematical calculations required in the HG examinations is cause for concern. For example, 2006 DoE HG Paper 2, Question 3.2.5 required students to read two values from a table and to subtract one number from the other number for two marks.[199] Given that such a mathematical skill is very elementary, is this the level required at SC level?

It has been argued that free-response answers, because they lower the probability of students guessing correctly, might be more difficult for students than choosing the correct answer questions (Nitko & Brookhart, 2007). This argument will be addressed in Section 6.2.1.1 on student performance.

6.1.2 Content standards

If a construct-driven approach to assessment were used, the "meaning of the construct [complex of knowledge, skills, or other attributes] guides the selection or construction of relevant tasks [behaviours or performances which reveal those constructs, and the tasks or situations which elicit those behaviors] as well as the rational development of scoring criteria" (Messick, 1994, pp. 16, 22). The final SC Biology examination mark was used to classify students at different levels of competence or mastery based on their performance in tasks set around the construct of content standards. In this study, content standards are defined as what students should know (topics) and are able to do with what they know (cognitive demand) (Section 3.7.2). "[A]rticulating clear content specifications [specified distinct topics] [for tests should be] rooted in the uniqueness of the content area and of understanding the psychology of learning" (Webb, 2006, p. 155). While the CBS explicated the facts, concepts and processes that should be learned at the SC level and stated, via the objectives and approaches to the syllabus, the different cognitive levels at which the topics should be taught, learned (Chapter 4, Tables 4.3 and 4.4) and examined (Chapter 2, Table 2.8), no rationale was given for the selection of topics or of the organization of these topics, between the three years of study, Grade 10 to 12.[200] The CBS also did not state reasons for the specified proportions of HOCS to be tested in the SC examinations (Chapter 2, Table 2.8).

This section presents the analyses of the content standards of the SC examinations by topic, by performance expectations, and lastly, as content, understood as paired combinations of topics

[199] Examples from school leaving Biology examinations in other countries showed questions which required students to perform more complex calculations (Gandal, 1994) that those observed in this study.

[200] The topics and the sub-topics comprising the three years of Biology study which culminated in the SC Biology examination were cross-referenced against the TIMSS Biology curriculum framework in Table 2.7 (Chapter 2).

and performance expectations. Similarity of content between question papers is then explored using a similarity index. A consideration of the relationship between the breadth and depth of content across the examination question papers completes this section.

6.1.2.1 *Topics* (Appendices 6.29 - 6.37)

Deciding what subject knowledge matters most in a domain has been described as both "tedious" and "painstaking" but it is crucial for assessment as it should reflect "disciplinary understanding" (Wineburg, 1997, p. 260) in regard to what is important to be learned. In the absence of a rationale for why particular topics were chosen and their relative emphases within the South African SC Biology examinations, one cannot interrogate the biological or educational reasons for choices of topics made by the examiners. This section describes the topics and their relative emphases (defined as the mean percentage weighting of marks within a paper) in the SC Biology examinations and compares them to the topics, and the topic emphases in the high-school exit examinations in some other countries (Tamir, 1996; Valverde, 2005).

In South Africa prior to 2001, the CBS did not require specific emphasis on the various topics in the SC Biology examination (Chapter 2, Table 2.8). Consequently, different examiners emphasized different topics both within and between years (Appendices 6.29 - 6.35) (Table 6.17). In 2001, when the SC examination for government schools was instituted, the modified CBS prescribed the same relative weightings for each topic in both HG and SG examinations, and thereby minimized variation in the subsequent national DoE HG and SG examinations (Appendices 6.36 & 6.37).[201,202] prior to 2001, the collective topic emphases from 1994 to 2000 were similar to the topic emphases written into the modified CBS, in which plant water relations and aspects of human physiology were the most emphasized (Table 6.17).

201 The author was present at a meeting convened by the national DoE early in 2001 to discuss the implementation of a national SC Biology examinations for government schools. At this meeting representatives from each of the provincial examining bodies (generally examiners from the previous years) voted to use, for the 2001 SC Biology examination onwards, the model used by KwaZulu-Natal in the 2000 SC Biology examination. There is no documented evidence for the rationale underpinning the weighting of topics that were selected for examination, or for why the previous single examination question paper was divided into two separate papers written on different days in the KwaZulu-Natal 2000 SC Biology examination (S, Chetty, personal communication, 25 November, 2008).

202 There was some variation between the weighting of the topics as calculated in this study relative to that of the modified CBS policy, because some components of knowledge appeared in more than one topic. In this study the author recorded such an item according to the topic in which it first appeared as listed in the CBS. Examiners might have differently classified any such topic so as to balance the weightings required by the modified CBS. Documentation of examiners classifications according to topic were available only for just two of the SC Biology examinations used in this study.

Table 6.17 Mean percentage weightings for topics in the 1994-2007 combined national DoE and IEB SC Biology HG and SG examinations.

Examination Body	Years	Bio-chemistry	Photo-synthesis	Respiration	Human nutrition	Human gaseous exchange	Population dynamics	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermo-regulation	Human circu-lation	Osmo-regulation	Outside syllabus
All	1994-2000	**12.9** (5.5)	7.8 (4.5)	6.9 (4.5)	**10.6** (5.1)	7.5 (4.9)	8.3 (5.1)	1.3 (1.6)	**12.1** (6.1)	7.9 (4.2)	**16.3** (6.3)	6.0 (4.9)	1.2 (3.7)	1.0 (1.3)	0.2 (1.7)
National	1994-2000	**12.7** (5.1)	8.2 (4.4)	7.4 (4.4)	**10.5** (4.9)	6.7 (4.2)	9.3 (4.4)	1.5 (1.6)	**12.5** (6.1)	7.6 (4.0)	**16.6** (6.3)	6.0 (4.7)		1.1 (1.4)	0.1 (0.4)
IEB	1994-2000	**14.0** (.37)	5.9 (4.4)	3.7 (3.9)	**10.8** (6.0)	**12.3** (5.8)	2.6 (5.2)	0.2 (0.6)	10.0 (6.2)	10.0 (5.0)	**14.3** (6.5)	6.3 (6.2)	8.6 (5.8)	0.3 (0.7)	1.0 (4.3)
National policy	2001-2007	9	7	6	**11**	8	9	2	**15**	**11**	**17**	5			
National HG	2001-2007	8.9 (2.5)	7.0 (1.4)	5.4 (1.4)	**10.0** (3.2)	9.7 (2.6)	9.0 (0.3)	2.0 (0.4)	**15.0** (0.5)	**10.3** (1.1)	**17.5** (1.8)	5.2 (0.8)			
National SG	2001-2007	**11.1** (2.2)	7.3 (1.2)	5.6 (1.3)	9.8 (2.7)	7.6 (1.3)	8.6 (0.7)	1.9 (0.6)	**14.4** (1.3)	**11.1** (0.5)	**16.9** (2.1)	5.7 (1.4)			
IEB HG	2001-2007	9.7 (3.0)	6.2 (7.6)	5.1 (2.6)	**14.1** (7.7)	7.7 (3.5)	7.2 (7.9)		**11.2** (5.6)	**11.5** (5.2)	**17.3** (6.9)	6.3 (2.6)	3.6 (4.5)		
IEB SG	2001-2007	**10.4** (5.2)	7.0 (3.7)	2.8 (3.2)	**11.4** (6.7)	9.3 (6.7)	5.5 (5.1)		**11.7** (5.3)	**12.2** (5.4)	**18.5** (4.7)	6.9 (3.7)	4.3 (5.7)		

Note: 1. National papers are the total obtained by combining data for topics across two papers written on different days.
2. Bold typeface indicates percentages greater than 10%.
3. Standard deviations are shown in brackets below the mean values.
4. IEB syllabus 1996 to 2003 ï human circulation replaces population dynamics.
5. *Amoeba* and *Lumbricus* ï included in human excretion, national DoE syllabus 1994 to 2003.

Recommended teaching time was also proportionally more for these two topics than for the other topics (Chapter 2, Table 2.5). Water relations was considered to be difficult to teach and to learn (T. Isaac, personal communication, July 16, 2010), and therefore one would have expected that more time be devoted to teaching this topic, but not necessarily that it would have been emphasized as much in the final examination. Osmoregulation in *Amoeba* and in the earthworm were removed as sub-topics of human excretion from the 2001 national DoE examinations. After 2000, the IEB HG examinations continued to show variation in the relative emphases of the different topics within both the HG and SG examinations of different years (Table 6.17). The IEB replaced population dynamics with human circulation in the 1996 to 2003 examinations,[203] and taught and examined population dynamics in Grade 10 (government schools taught and examined human circulation in Grade 10). The IEB re-arrangement made sense, as population dynamics fits biologically with ecology which was part of the Grade 10 syllabus, and human circulation joined other aspects of human physiology in the SC Grade 12 syllabus. Post-2000, the IEB emphasized all topics similarly in their HG and SG examinations written in the same year (Table 6.17).

Tamir (1996) and Valverde (2005) compared the emphases of various different topics in the high school leaving Biology examinations of a number of countries. While Tamir (1996) and Valverde (2005) both used adaptations of the TIMSS framework in their comparisons, each quantified the emphases in different ways making direct comparisons of their work difficult.[204] Tamir (1996, p. 73) provided quantitative details of the five topics emphasized in each of the high school leaving Biology examinations of England and Wales, France, Germany, Israel, Japan and the USA, and Valverde (2005, pp. 41, 42) recorded the core topics in the examinations of the Middle East North African (MENA) countries: Egypt, Iran, Jordan, Lebanon, Morocco, Tunisia and France. Both Tamir (1996) and Valverde (2005) observed considerable variation in the emphases on topics in the examinations. Tamir (1996, p. 72) speculated that "a lack of commonality in topics may in part arise from the nature of the field [biology]" or contextual differences between countries. For example, German examinations emphasized animal behaviour perhaps because many prominent ethologists were German (Tamir, 1996).

At most a superficial comparison of the emphases of the South African SC Biology examinations with those of other countries (Tamir, 1996; Valverde 2005) could be made

[203] The IEB was required to revert to its pre-1996 SC Biology topics, so that their papers covered the same topics as the national DoE SC examinations for comparability purposes (T. Isaac, personal communication, December 10, 2004).

[204] The intended curriculum of the South African Grades 10, 11 and 12 (SC) was mapped against the TIMSS framework in Table 2.7 (Chapter 2).

here.[205] Tamir (1996) expressed an expectation that energy handling (i.e., photosynthesis, respiration), variation and inheritance, cell biology, sensing and responding and biological processes (i.e., metabolism, protein synthesis, cell water relations) would be emphasized in school-leaving examinations, because they were among the core aspects of biology. While these topics were not all covered in the SC senior school examinations, they were all stipulated as part of the intended and assessed curricula of the preceding years, Grades 10 and 11 (Chapter 2, Table 2.7). Three of the topics, that is, photosynthesis, respiration and water relations, identified by Tamir (1996) as important, were examined in most of the SC examination analyzed but were also absent from some of the examinations or appeared in optional questions. As these three topics all addressed abstract concepts and require some knowledge of biochemistry to be understood, it was expected that their presence or absence might have impacted on student performance in different examinations and hence their comparability with respect to standards.[206] This expectation will be discussed in Section 6.2.2.1.

Evolution, which is considered the cornerstone to understanding present and past life-forms (Dobzhansky, 1973) and "the most important concept in modern biology, a concept essential to understanding key aspects of living things" (National Academy of Sciences, 1998, p. viii), was only present in the school-leaving examinations of Jordan, Tunisia and Germany (Bavaria) (Tamir, 1996; Valverde 2005). While absent from the SC Biology examinations analyzed here, the importance of evolution to the high school curriculum was recognized in the new Life Sciences curriculum examined for the first time in the 2008 NSC examinations.

In this study, only two questions were considered to be on topics outside of the CBS SC examination syllabus. One of these questions (Figure 6.5 [A]) acknowledged within the question that it did not directly access content prescribed by the CBS and required students to draw on concepts that would have been learned in earlier grades about the biochemistry of organic molecules, which had been part of the CBS prescribed content for those grades. The second question, optional and philosophical, required students to draw on their knowledge and understanding about Biology at the end of their high school careers in the context of a quote made by Ernst Mayr[207] (Figure 6.5 [B]). One alternative essay to this second question required students to explore the relationship between human excretion and a range of homeostatic process in humans, and another to compare the human eye with a camera.

[205] In order to make meaningful comparisons between this study about South Africa and that of Tamir (1996) and Valverde (2005) the author would have needed more quantitative data than appeared in these two publications.

[206] "For students to be well-prepared for future study in biology, they should understand both ends [abstract molecules to larger scale ecosystems]" (Tamir, 1996, p. 81).

[207] Ernst Mayr is considered one of the most famous evolutionary biologists of the 20th century.

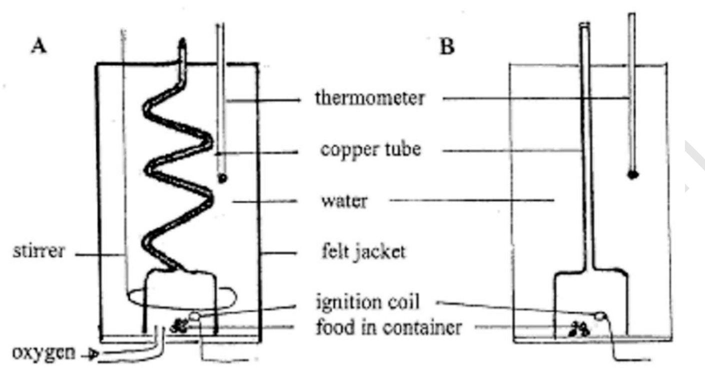
A. Natal Education Department 1994 HG

Question 4.4

This apparatus does not directly relate to the content of the syllabus. The purpose of the question is to test your ability to:

- observe
- apply biological principles in a new situation

The diagrams below show apparatus which can be used to find the energy content of food. The food is burnt by an ignition coil. The increase in temperature of the surrounding water indicates the amount of energy present in the food.



4.4.1 State <u>three</u> differences between apparatus A and [apparatus] B. (3)
4.4.2 Explain the reason for these differences stated in 4.4.1. (6)

B. IEB 1998 HG

The essay question below is not based on any one area of the syllabus and particular emphasis will be place on your ability to critici[z]e and argue effectively as well as the broad knowledge of biology and the scientific outlook that you show in your answer.

Question 8

Ernst Mayr in his book ñThis is Biologyò (Belkap Press 1997) asks the question ñIs biology, like physics and chemistry, a science?ò He answers this question by referring to eight criteria for a science. These include:

a. A science must be based on data collected by observation or experiment.
b. Data must be collected to answer questions: and observations must be made to strengthen or refute conjectures.
c. Objective methods must be employed to minimize bias.
d. Hypotheses must be consistent with the observations and compatible with what is known.
e. All hypotheses must be tested and compared with competing hypotheses.
f. Generali[z]ations must be universally valid; and unique events must be explainable without reference to supernatural causes.
g. A fact of discovery must be fully accepted only if repeatedly confirmed by other investigations.
h. There should be a steady improvement of scientific theories by replacing faulty ones and by solving previously puzzling problems.

Write an essay in which you discuss how far biology at school level can be considered a science in view of THREE of these criteria. (60)

Figure 6.5 Questions classified in this study as outside of the CBS.

Although both these alternate essay questions required more than the simple recall of knowledge, the philosophical question was probably the most difficult to answer. One might hypothesize that a question such as the essay question in Figure 6.5 (B) might have been chosen by the more able students and that both these questions (Figure 6.5 [A and B]) might have distinguished between more competent and less competent students. Unfortunately, no records were available as to how students performed in either of these questions nor could memoranda for these questions be sourced to establish the type of answers which examiners expected from these answers.[208]

Comparing questions is necessary for determining matters of equivalence which are important with regard to fairness to students. The two questions discussed above which are outside of the syllabus flag how necessary memoranda are when analyzing questions. What was important was that both these questions gave students the opportunity to demonstrate coherence in their understanding of biology which is prized as an indicator of high-quality standards (Schmidt, Houang, & Cogan, 2002; Schmidt et al., 2005). These questions also demonstrated how the narrowing of topics examined so often associated with the use of essays in biology examinations (Tamir, 1996) could be obviated.

6.1.2.2 *Performance expectations* (Appendices 6.38 - 6.46)

Higher order cognitive demand, abbreviated to HOCS in this study, is increasingly implicated as being important in how individuals are taught and what they learn (e.g., Alberts, 2009, 2010; NRC in the USA have emphasized HOCS (Porter, McMaken, Hwang, & Yang, 2011). However, international benchmarking of the core content standards in two USA subjects, Mathematics and English Arts Language and Reading, with the content standards of countries considered to be top-achievers in terms of student performance showed the reverse, that is, top-achieving nations placed greater emphasis on students performing procedures than on demonstrating HOCS (Porter et al., 2011a). While Porter et al. (2011a) expressed surprise at this finding, it may simply arise because the framework they used lacked an explicit or articulated connection between HOCS and LOCS. Top-achieving nations which placed curricular emphases on the performance of procedures may thus not necessarily have been de-emphasizing HOCS—rather they may have been implicitly recognizing that LOCS are a necessary pre-requisite for successful performance in HOCS. If this pre-condition is indeed recognized, the call for the active learning of HOCS in introductory biology courses (e.g., Haak, et al., 2011) and more emphasis on HOCS in curricula (e.g., Schmidt et al., 2005) should

[208] The author asked individuals who had been involved with these examinations and no records or memories of either of these questions were found.

be expanded to include the explicit promotion of active learning of *both* LOCS and HOCS in all Biology courses, at school- and college-level. Such expanded learning would then be in line with a contemporary constructivist teaching and learning approach (e.g., Lord, 1997). The author's different interpretation of the Porter et al. (2011a) and Haak et al. (2011) findings result from the different ways in which LOCS and HOCS are viewed across studies, and illustrates how difficult it is it make comparisons of cognitive demand between studies.

The difficulty of comparative studies of cognitive demand is further illustrated in two studies. In the first, Tamir (1996, p. 81) in his comparison of the performance expectations in high school leaving Biology examination in school leaving Biology examinations of England and Wales, France, Germany, Israel, Japan and the USA found that the examinations of all of the countries, other than France, involved mainly questions which required the recall of information "however sophisticated the concepts involved may be". Tamir (1996) classified these questions as Understanding, the lowest level of the TIMSS framework that he used.[209] Using a similar tool to that of Tamir (1996) to benchmark the performance expectations in high school leaving Biology examination in six different MENA countries against those of France, Valverde (2005)[210] found similar emphases on Understanding in the MENA and French examinations. However, Valverde (2005, p.46) interpreted a question classified as understanding information to refer "to an examinee's ability to demonstrate understanding of vocabulary, facts, equations, and simple concepts. These test tasks thus provided opportunities for students to define, describe, and name simple concepts". From these descriptions it was difficult to separate the emphases of recall from understanding in the analyses of performance expectations presented in the comparisons made by Tamir (1996) and Valverde (2005). This challenge highlights the importance of explicitness of classifications by cognitive demand, if comparisons are to be made. The danger of mixing the use of contrasting measures of cognitive demand within comparisons was discussed in Chapter 4. In this study, questions which did not require students to demonstrate evidence that they understood facts, concepts, processes or routine procedures were classified as either Memorize or Perform-Routine- Procedures. In consequence, the results of this study cannot be directly compared to those of Tamir (1996) and Valverde (2005).

That HOCS are not easy to measure, or not easy to measure in reliable and cost-effective ways, should not undermine the necessity or the urgency of finding ways to address these challenges (Silva, 2009). Given the small extent of HOCS in South African SC Biology examination

---

209 TIMSS Curriculum Frameworks (Robitaille et al., 1993).

210 Valverde (2005) used slightly changed descriptions of the TIMSS categories and did not report on the TIMSS category: using tools, routine procedures and science processes.

question papers, we may ask whether students are being given sufficient opportunity to demonstrate their HOCS abilities?

In South Africa the CBS objectives (Chapter 4, Table 4.3) and the approach to the syllabus (Chapter 4, Table 4.4) prior to 2001, required that students access the content stipulated in the CBS at all levels of cognitive demand (Chapter 2, Figure 2.4). The CBS also required that specified percentages of the SC examinations be devoted to testing higher intellectual skills (Table 2.8), but gave no indication as to exactly what was meant by higher intellectual skills (Figure 4.3 [A]). In 2001, the modified CBS, used by the national DoE, proposed an adapted version of BTEO to define what was meant by higher intellectual skills (Chapter 4, Figure 4.3 [B]). This version was different to the adapted version of BTEO used by the IEB (Chapter 4, Figure 4.3 [C]). Neither the DoE nor the IEB provided an accompanying rationale for their particular interpretations of what higher order skills are. Consequently, in this study it was difficult to consistently compare examinations within and between years. The PET was therefore developed as an alternative for use in this study (Chapter 4).

The PET groups performance expectations, which can be recognized in examination questions, according to different categories of cognitive demand. This approach argued for the recognition of HOCS as being those categories which require students to make connections between content that is not explicitly stated in the CBS or the modified CBS. The author acknowledges that the use of different instruments to measure cognitive demand tells a different story about cognitive demand, but argued that the use of PET is justified in this study given the absence of an explicit policy instrument or means to recognize cognitive demand (Chapter 4).

Thirty years ago Champagne and Klopfer (1981) demonstrated that it was not enough to assess the scientific facts and concepts (called content by these authors) that students had learned, and that it was necessary to provide students with the opportunities to demonstrate how they could use what they had learned to solve problems. Problem-solving is a HOCS in this thesis.

In the question papers analyzed in this study, none of the HG examinations approached the target 40% HOCS and none of the SG examinations approached the 20-25% HOCS required by the CBS policy or the modified CBS policy. As expected because of policy directives within each of the years 2001 to 2007, the HG question papers of both the DoE and the IEB had a greater number of HOCS questions than did the SG questions papers of the same examination session. Application rather than Analysis questions contributed mostly to the HOCS questions which appeared in the question papers analyzed in this study. Apply type questions require students to show evidence that they can make connections between what is learned and/or given in the question (Analyze) and a new context.

A ministerial committee appointed by the DoE to investigate the SC examinations quoted a Biology HG examiner as saying: "The way of mixing questions from different chapters is of disadvantage to candidates. Questions should be set according to chapter … this [sequence] will enable candidates to target specific questions that they can score most on, rather than confusing them" (DoE, 1998, p. 11). Therefore, this attitude amongst examiners who set the question papers and the fact that PET considers Analyze questions to require students to make connections not required by the syllabus but in familiar contexts (like different chapters), make the lack of Analysis questions is not surprising. What is surprising, is that this approach persisted given that the ministerial committee reported that "[t]his [reliance of rote learning in SC examinations] must be addressed as soon as possible" (DoE, 1998, p.11). The 2001 reworking of the CBS by the DoE did not explicitly address these concerns.[211] Making connections between topics increases meaning and retention (Sousa, 2006).

Schmidt et al. (2002, p.9) regarded content standards as coherent if they were "articulated over time [within and between grades] as a sequence of topics and performances that are logical and reflect, where appropriate, the sequential or hierarchical nature of the disciplinary content from which the subject matter derives". Schmidt et al. (2005, p.554). considered "coherence is one of the most critical, if not the single most important, defining elements of high-quality standards". The way that the CBS was organized, with little explicit requirement that students make connections between content across topics, or between topics studied in earlier years, probably promoted the lack of connectedness in content observed in the SC examinations. In 2004, osmoregulation in *Amoeba* and *Lumbricus* was removed as a topic from the CBS syllabus,[212] decreasing opportunities for examiners to ask questions which would have required

211 In South Africa, there is continued confusion between what constitutes analysis type questions and application type questions in the NSC current Life Sciences examinations. Reasons given for the downward adjustment of the 2011 NSC Life Sciences marks for certification was that "[t]he [question] papers were less demanding and didn't contain sufficient application questions" (Umalusi, 2011, no page number given). The instrument used by the examiners when they set these question papers indicated that the HOCS questions were *all* from the category application, which was acceptable given that the policy did not unambiguously indicate that HOCS should come from both application and other higher cognitive categories such as analysis (Crowe, 2010). There is not a common understanding of the meaning of 'application' as a level of cognitive demand in South Africa, or globally (Chapter 4).

212 Previously osmoregulation in *Amoeba* and *Lumbricus* had been included in the topic on human excretion. The author could find no documentation about the decision made to not examine osmoregulation in *Amoeba* and *Lumbricus* in the SC examinations. One of the Umalusi external moderators involved in making this decision explained that the argument given for the removal of these topics was that these topics were a part of the Grade 11 syllabus and would therefore have

students to demonstrate conceptual understanding of the role of osmoregulation in all organisms, and not just in humans. Despite the lack of coherence opportunities for examiners because The CBS and modified CBS policies required that only Grade 12 topics be examined in the SC examinations, Analysis questions could have been examined in several other ways, such as combining unlearned information given in the question itself with learned knowledge, or philosophical questions as discussed in Section 6.1.2.1 above.

Analysis type questions appeared in the HG question papers each year from 1994 to 2000 for the CED and the WCED (Table 6.18) but did not appear in any of the DoE HG and SG question papers post-2000 (Table 6.19). A noticeable exception was the DoE 2005 HG examination where both question papers had an unusually high number of questions requiring students to combine information located in a combination of non-text representations (therefore classified as Analysis questions), for example, data from a table and a graph or from a diagram and a graph. The IEB HG question papers for all years, except 2005 which had more Application questions, included Analysis questions (Table 6.20). Therefore, the 2005 HG examinations varied with respect to the percentage of Analysis questions which they contained. A similar dearth of Analysis questions in both the DoE and IEB SG examinations (Tables 6.19 & 6.21) was expected, and observed, given that policy was not clear about what HOCS were. The Application questions in SG question papers were almost always of the type of question which did not involve the use of analysis skills (i.e., Chapter 4, ii in Figure 4.7). Two questions need to be asked: how was equivalence between different examinations with different percentages of HOCS, and those examinations of previous years established? And, how can equivalence be addressed without a common understanding of what both LOCS and HOCS are?

Given that different instruments used to measure cognitive demand describe cognitive demand in different ways, that what constitutes HOCs policy differs between examining bodies (Chapter 4) and the absence of records of examiners' cognitive demand analyses for the different SC Biology examinations, it might be considered unfair to make judgments about comparable levels of cognitive demand between examinations, or their lack of HOCS. However, the results reported here do indicate a point of urgency: that is, the necessity for policy to be explicit about what different levels of cognitive demand mean (especially what constitutes HOCS), and how cognitive demand might be consistently recognized, and

been examined for promotion purposes at the end of Grade 11 (T. Isaac, personal communication, August 16, 2011).

Table 6.18 Percentage weightings of categories of cognitive demand in the 1994, 1995 CED and 1996-2000 WCED SC Biology HG question papers.

Year	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
1994	1	400	53	6	24	9	9
1994	2	400	66	6	13	9	7
1995	1	400	62	10	12	5	11
1995	2	400	60	10	13	5	12
1996	1	300	75	3	12	8	3
1996	2	300	75	3	16	2	4
1997	1	300	83	4	6	2	5
1997	2	300	85	4	4	2	5
1998	1	300	71	4	11	5	10
1998	2	300	77	1	9	3	10
1999	1	300	63	8	15	7	6
1999	2	300	64	8	14	7	6
2000	1	300	59	4	18	7	12
2000	2	300	63	3	17	7	10

Table 6.19 Percentage weightings of the different categories of cognitive demand in the 2001-2007 national DoE SC Biology HG and SG question papers.

Year, grade	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
HG							
2001	1	400	52	10	24	1	14
2002	1	400	46	17	20	0	18
2003	1	400	48	11	25	0	17
2004	1	400	41	11	23	1	24
2005	1	400	41	11	13	14	21
2006	1	400	39	13	26	6	16
2007	1	400	48	9	21	0	23
SG							
2001	1	300	67	10	20	0	4
2002	1	300	72	4	19	0	5
2003	1	300	64	10	18	0	8
2004	1	300	60	7	26	1	7
2005	1	300	65	6	14	1	14
2006	1	300	65	6	23	0	6
2007	1	300	62	10	18	0	9

Table 6.20 Percentage weightings of the different categories of cognitive demand in the 1994-2000 IEB SC Biology HG question papers.

Year	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
1994	1, 2	400	32	4	40	4	21
1995	1, 2	400	45	6	24	1	25
1996	1, 2 5, 6	400	49	9	22	3	18
1997	1, 2 3, 4	400	61	7	11	10	11
1997	5, 6 7, 8	400	58	10	11	10	11
1998	1, 2	320	51	5	23	12	9
1998	3	320	44	5	23	3	26
1998	4,5	320	54	5	19	12	9
1998	6	320	47	5	19	3	26
1999	1, 2	320	48	4	32	7	9
1999	3, 4	320	48	4	29	7	12
2000	1, 2 3, 4	320	46	6	30	8	10
2000	5, 6 7, 8	320	49	6	27	8	10

Table 6.21 Percentage weightings of the different categories of cognitive demand in the 2001-2007 IEB SC Biology HG and SG question papers.

Year, grade	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
HG							
2001	1, 2	300	45	7	21	15	12
2002	1, 2	400	53	8	25	8	6
2003	1, 2	300	42	7	28	13	10
2004	1, 2	300	43	7	26	4	20
2005	1, 2	300	42	7	26	0	25
2006	1, 2	300	40	10	26	3	21
2007	1, 2	300	42	12	22	0	23
SG							
2001	1, 2	225	64	5	27	3	2
2002	1, 2	300	63	5	16	10	7
2003	1, 2	225	77	4	16	0	3
2004	1, 2	225	75	4	19	0	2
2005	1, 2	225	82	4	10	0	4
2006	1, 2	225	63	4	15	0	18
2007	1, 2	225	73	3	12	1	11

nominally described in both teaching and assessment.[213] Since recent research indicated a lack of HOCS in the teaching, learning and assessing of sciences like Biology at university level (Lord & Baviskar, 2007; Zheng et al., 2008), developing a common understanding of HOCS should not be restricted to the school level of education systems.

6.1.2.3 *Content (topics and performance expectations)*

Conceptually, topics and performance expectations in combination describe the content standards of examination question papers (Chapter 3). Porter (2002) described and explained the value, especially for comparative purposes, of visually displaying the two-dimensional framework defining content at the intersections of topics and categories of cognitive demand, as content maps. The reliability of content analyses using this framework has been demonstrated (Porter, 2002; Porter et al., 2009). Visual displays of the content standards of HG examinations for the CED-WCED-DoE (examples of government administered examining bodies) and the IEB (an example of a non-government administered examining body) are used here to demonstrate the kinds of information that can be obtained by inspection of a time series of these diagrams (1994 to 2007) (Figures 6.6 & 6.7). The topic dimension is constituted by the thirteen discrete topics recognized in this study (Chapter 5, Table 5.5) and the cognitive demand dimension is constituted by the five groups of performance expectations (Chapter 4, Figure 4.6). Thus, for this study, there are 65 distinct types of content recognized.[214] Some examinations supported multiple profiles if they offered a choice of questions, each of which addressed different content standards. Content standards represented in Figures 6.6 and 6.7 differ from visual representations given in Porter (e.g., 2002, 2011) because absolute raw cell scores, rather than cell proportions, are used here. This approach clearly exhibits that different SC Biology question papers carried more or less total marks than did other question papers[215,216] and the content maps are not converted into t content maps. Some examinations had multiple

---

213 Britton et al. (1996a, p. 48) argued that as the application of scientific knowledge and principles is core to scientific activities, it should be expected that a "substantial proportion of examinations" be "devoted to" performances demonstrating scientific competence. While these authors gave no absolute value for what 'substantial' might mean in these contexts, they lauded the Swedish science high school leaving examinations which had 68% of their questions addressing theorizing, analyzing and solving problems, a category of the TIMSS framework borrowed for their comparative international study (Britton, 1996a).

214 This set of 65 would be considered a relatively coarse-grained analysis because of the small number of content cells (Porter, 2002). A finer grain could have been obtained by sub-dividing the topics into sub-topics and/or dividing the performance expectations into sub-categories of performance expectations. The author tried a finer grain in the initial analyses but because of the wide range of different sub-topics and individual performance expectations. It was difficult to make meaningful comparisons of content across question papers at this finer level.

215 Topics and performance expectations are each presented elsewhere as percentages of marks of each question paper (Tables 6.17 - 6.21).

216 The number of marks can influence the difficulty of an examination (Section 6.3) and the equivalence between question papers (Section 6.4).

profiles because they offered a choice of questions each of which addressed different content standards, in which case intra-year variation in content was highlighted by shading the relevant content cells. Figures 6.6 and 6.7 highlight the differences in content between various admissible profiles within the examination of one year for one examining body.

A choice of questions was not offered in the DoE examinations (2001 ─ 2007) (Table 6.3) but choices of questions in the CED and the WCED examinations (1994 ─ 2000) (Table 6.2) resulted in two unique content profiles for each year (Figure 6.6). Profile 1 and Profile 2 of these examinations show that differences between profiles were not consistent between years. For example, within one year the CED 1994 HG Profile 1 placed less emphasis on Memorize-Photosynthesis and Perform-Routine-Procedures-Photosynthesis, and more emphasis on Perform-Routine-Procedures-Thermoregulation, Explain-Thermoregulation and Apply-Thermoregulation, than did Profile 2 (Figure 6.6). The IEB offered a choice of questions in all their HG examination question papers from 1994 to 2007 (Tables 6.4 & 6.5). This set of choices resulted in, for example, eight unique content profiles in 1997 (Figure 6.7), each of which had very different content emphases to the others. No other intra-year, inter-year or inter-examining bodies comparisons showed the same large content differences described in these two examples of time series content maps (Figures 6.6 & 6.7). None of the years suggested the same pattern of distributions of content emphases. The DoE post-2000 policies ensured that eleven of the topics were examined every year thereafter in the DoE examinations (Figure 6.6). Without a policy which required all topics to be examined each year, the IEB did not examine all topics each year (Figure 6.7). Figures 6.6 and 6.7 also indicate that in some years more content categories was examined, often at LOCS levels, than in other years – what Schmidt et al. (2005, p. 555) might have described as a "mile-wide[,] inch-deep" assessed curriculum, where a large number of topics are assessed with little depth. The relationship between topic coverage and depth coverage in the SC examinations analyzed here will be discussed in section 6.1.2.3 below.

Given the variation in content leads one must ask: how was equivalence in content between these question papers recognized? Or, how much similarity between content is sufficient for equivalence to be assumed between question papers? The next section demonstrates one way that similarity of examinations can be explored.

1994 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	31		12		8
Photosynthesis	3				
Respiration					
Human nutrition	49.5			28	
Human gaseous exchange	12			6	4
Population dynamics	13		12	3	2
Plant hormones	3				
Plant water relations	26	6	20		13
Human excretion	4.5	8	6		
Human co-ordination	54.5				2
Thermo-regulation	14	8	44		6
Human circulation					
Osmoregulation	1.5				
Out of syllabus					

1994 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	31		12		8
Photosynthesis	55	8			
Respiration					
Human nutrition	49.5			28	
Human gaseous exchange	12			6	4
Population dynamics	13		12	3	2
Plant hormones	3				
Plant water relations	26	6	20		13
Human excretion	4.5	8	6		
Human co-ordination	54.5				2
Thermo-regulation	12				
Human circulation					
Osmoregulation	1.5				
Out of syllabus					

1995 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	47	4	3	5	11
Photosynthesis	11	4		3.5	
Respiration	13			1.5	2
Human nutrition	16	4			4
Human gaseous exchange	13	6	4	4	
Population dynamics	21	8	18	6	5
Plant hormones					
Plant water relations	21	4	3		18
Human excretion	11	8	12		
Human co-ordination	35		8		3
Thermo-regulation	61				2
Human circulation					
Osmoregulation					
Out of syllabus					

1995 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	47	4	3	5	11
Photosynthesis	11	4		3.5	
Respiration	13			1.5	2
Human nutrition	64	4	2		14
Human gaseous exchange	13	6	4	4	
Population dynamics	21	8	18	6	5
Plant hormones					
Plant water relations	21	4	3		18
Human excretion	11	8	12		
Human co-ordination	35		8		3
Thermo-regulation	3				
Human circulation					
Osmoregulation					
Out of syllabus					

1996 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	37	2			
Photosynthesis	11		2.5	10	
Respiration	13		2.5	10	
Human nutrition	33		15		
Human gaseous exchange	28		6		
Population dynamics	18	2	10	1	2
Plant hormones	2				
Plant water relations	21	4		1	6
Human excretion	8				
Human co-ordination	40				
Thermo-regulation	10			2	
Human circulation					
Osmoregulation	3				
Out of syllabus					

Figure 6.6 continued on next page

1996 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	37	2			
Photosynthesis				1	
Respiration	2			1	
Human nutrition	33		15		
Human gaseous exchange	28		6		
Population dynamics	18	2	10	1	2
Plant hormones	2				
Plant water relations	21	4		1	6
Human excretion	8				
Human co-ordination	40				
Thermo-regulation	34	4	14	2	3
Human circulation					
Osmoregulation	3				
Out of syllabus					

1997 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	26	4			9
Photosynthesis	16				
Respiration	2				
Human nutrition	28		6	6	
Human gaseous exchange	5	2	6		
Population dynamics	33	5			
Plant hormones	2				
Plant water relations	42		5		
Human excretion	32				
Human co-ordination	45				7
Thermo-regulation	15				
Human circulation					
Osmoregulation	4				
Out of syllabus					

1997 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	26	4			9
Photosynthesis	16				
Respiration	17				
Human nutrition	38		6	6	
Human gaseous exchange	25	2	6		
Population dynamics	33	5			
Plant hormones	2				
Plant water relations	2				
Human excretion	32				
Human co-ordination	45				7
Thermo-regulation	15				
Human circulation					
Osmoregulation	4				
Out of syllabus					

1998 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	34	2			6
Photosynthesis	26		3		
Respiration	8		2		
Human nutrition	25		3	3	4
Human gaseous exchange					
Population dynamics	10	11	9		
Plant hormones	2				
Plant water relations	18		3	6	9
Human excretion	24		7		6
Human co-ordination	57		5	1	4
Thermo-regulation	6			4	
Human circulation					
Osmoregulation	2				
Out of syllabus					

1998 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	34	2			6
Photosynthesis	14		3		
Respiration	8		2		
Human nutrition	25		3	3	4
Human gaseous exchange					
Population dynamics	1				
Plant hormones	2				
Plant water relations	18		3	6	9
Human excretion	24		7		6
Human co-ordination	97		10	1	4
Thermo-regulation	6				
Human circulation					
Osmoregulation	2				
Out of syllabus					

Figure 6.6 continued on next page

1999 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	19	18		4	
Photosynthesis	7		1		
Respiration	22		5		
Human nutrition	35	1	3	4	8
Human gaseous exchange	25		2		
Population dynamics	15	5	7		1
Plant hormones	2				
Plant water relations	28		10	10	2
Human excretion	3	1	7	1	6
Human co-ordination	30		9		2
Thermo-regulation	4				
Human circulation					
Osmoregulation			2	1	
Out of syllabus					

1999 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	32	18	2	4	
Photosynthesis	20				
Respiration	2		3		
Human nutrition	37	1	3	4	8
Human gaseous exchange	10				
Population dynamics	15	5	7		1
Plant hormones	7				
Plant water relations	33		10	10	2
Human excretion	3	1	7	1	6
Human co-ordination	30		9		2
Thermo-regulation	4				
Human circulation					
Osmoregulation			2	1	
Out of syllabus					

2000 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	25	1	5	5	1
Photosynthesis	6				
Respiration	8		7		2
Human nutrition	18	6		3	17
Human gaseous exchange	11	2	15		
Population dynamics	16	1	9		9
Plant hormones	1				
Plant water relations	35	1	10		5
Human excretion	19		4		
Human co-ordination	37		5	13	1
Thermo-regulation					
Human circulation					
Osmoregulation	2				
Out of syllabus					

2000 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	25	1	5	5	1
Photosynthesis	6				
Respiration	8		7		2
Human nutrition	18	6		3	17
Human gaseous exchange	11	2	15		
Population dynamics	16	1	9		9
Plant hormones	1				
Plant water relations	6				
Human excretion	19		4		
Human co-ordination	37		5	13	1
Thermo-regulation	40		5		
Human circulation					
Osmoregulation	2				
Out of syllabus					

Figure 6.6 continued on next page

2001 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	21		6		18
Photosynthesis	12.5		10	3	
Respiration	23.5				
Human nutrition	31		9		
Human gaseous exchange	20		10		
Population dynamics	3	8	16		9
Plant hormones	5				
Plant water relations	34	14	6		6
Human excretion	9		19		4
Human co-ordination	41		18	2	19
Thermo-regulation	6		17		
Osmoregulation					
Out of syllabus					

2002 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	26				5
Photosynthesis	20	4			14
Respiration	3	2	4		6
Human nutrition	13	5	18		4
Human gaseous exchange	18	7	7		6
Population dynamics	10	21	2		5
Plant hormones	8				
Plant water relations	23	18	18		4
Human excretion	14	8	6		16
Human co-ordination	40	2	10		8
Thermo-regulation	7		15		3
Osmoregulation					
Out of syllabus					

2003 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	18	4	9		16
Photosynthesis	2	11	8		5
Respiration	8	6	7		3
Human nutrition	23	2	6		6
Human gaseous exchange	13	4	7		8
Population dynamics	20		4		10
Plant hormones	8				
Plant water relations	33	8	12		7
Human excretion	17		24		3
Human co-ordination	33	7	20		8
Thermo-regulation	17		3		
Osmoregulation					
Out of syllabus					

2004 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	14	**3**	4		22
Photosynthesis	15		5		
Respiration	10		4		12
Human nutrition	16				
Human gaseous exchange	17		18		23
Population dynamics	12	15	10		
Plant hormones	9				
Plant water relations	18	23	11		5
Human excretion	12		9		18
Human co-ordination	36	2	28	5	9
Thermo-regulation	4	2	4		5
Osmoregulation					
Out of syllabus					

2005 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	11	**3**	2	1	18
Photosynthesis	10	8	3	3	3
Respiration	5	2	6	8	7
Human nutrition	22	4	4	8	5
Human gaseous exchange	27		3	2	
Population dynamics	21	5		1	8
Plant hormones	3			2	3
Plant water relations	13	13	9	8	17
Human excretion	16	8	10	3	7
Human co-ordination	31		7	13	17
Thermo-regulation	6		7	7	
Osmoregulation					
Out of syllabus					

Figure 6.6 continued on next page

	2006 Profile 1					2007 Profile 1				
	Memorize	Procedure	Explain	Analysis	Application	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	9				20	8	3			9
Photosynthesis	12	2	9	3.5	2	12.5		4.5		12
Respiration	8			3.5		20		2		2
Human nutrition	20	19			8	36	2	8		12
Human gaseous exchange	30	9	7		2	29		4		
Population dynamics	5	6	18		7	12	11	6		7
Plant hormones	8		2			8				
Plant water relations	26	4	18		10	26	13	12		11
Human excretion	7	12	18		5	20		18		5
Human co-ordination	32		20	8	7	17	6	18		26
Thermo-regulation			13	7	3	3		11		6
Human circulation										
Osmoregulation										
Out of syllabus										

Figure 6.6 Variation in the composition of content (raw scores) for 1994,1995 CED, 1996-2000 WCED and 2001-2007 national DoE Biology HG examinations. Shaded cells indicate differences between different profiles within a year.

1994 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	48		8		14
Photosynthesis	2		22	12	3
Respiration	9	2	6		16
Human nutrition	28	8	28		
Human gaseous exchange	2		6	4	6
Population dynamics	6	3	4		28
Plant hormones	2		5		
Plant water relations	2		20		
Human excretion	2				
Human co-ordination	19		59		
Thermo-regulation	6	4			16
Human circulation					
Osmoregulation					
Out of syllabus					

1994 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	48		8		14
Photosynthesis	2		22	12	3
Respiration	9	2	6		16
Human nutrition	4				
Human gaseous exchange	2		6	4	6
Population dynamics	6	3	4		28
Plant hormones	2		5		
Plant water relations	2		20		
Human excretion	26	8	28		
Human co-ordination	19		59		
Thermo-regulation	6	4			16
Human circulation					
Osmoregulation					
Out of syllabus					

1995 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	60	8	43		12
Photosynthesis	10				22
Respiration	12				
Human nutrition	17		4	1.5	
Human gaseous exchange	3	6	16		
Population dynamics	19		7		30
Plant hormones					
Plant water relations	17				23
Human excretion	7		15		10
Human co-ordination	34	8	6	1.5	
Thermo-regulation	2		3		3
Human circulation					
Osmoregulation					
Out of syllabus					

1995 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	36		15		12
Photosynthesis	10				22
Respiration	12				
Human nutrition	17		4	1.5	
Human gaseous exchange	3	6	16		
Population dynamics	19		7		30
Plant hormones					
Plant water relations	17				23
Human excretion	7		15		10
Human co-ordination	58	16	34	1.5	
Thermo-regulation	2		3		3
Human circulation					
Osmoregulation					
Out of syllabus					

1996 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	63				28
Photosynthesis	15	5	17.5		
Respiration					
Human nutrition	24			3	
Human gaseous exchange	25	4	25		
Population dynamics					
Plant hormones					
Plant water relations	15	5	17.5		
Human excretion	9	19			20
Human co-ordination	25			2	18
Thermo-regulation					4
Human circulation	20	4	26	6	
Osmoregulation					
Out of syllabus					

Figure 6.7 continued on next page

1996 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	63				28
Photosynthesis					
Respiration					
Human nutrition	24			3	
Human gaseous exchange	25	4	25		
Population dynamics					
Plant hormones					
Plant water relations					
Human excretion	9	19			20
Human co-ordination	25			2	18
Thermo-regulation	30	10	35		4
Human circulation	20	4	26	6	
Osmoregulation					
Out of syllabus					

1996 Profile 5

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	50				28
Photosynthesis	28	5	17.5		
Respiration					
Human nutrition	24			3	
Human gaseous exchange	25	4	25		
Population dynamics					
Plant hormones					
Plant water relations	15	5	17.5		
Human excretion	9	19			20
Human co-ordination	25			2	18
Thermo-regulation					4
Human circulation	20	4	26	6	
Osmoregulation					
Out of syllabus					

1996 Profile 6

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	50				28
Photosynthesis	13				
Respiration					
Human nutrition	24			3	
Human gaseous exchange	25	4	25		
Population dynamics					
Plant hormones					
Plant water relations					
Human excretion	9	19			20
Human co-ordination	25			2	18
Thermo-regulation	30	10	35		
Human circulation	20	4	26	6	
Osmoregulation					
Out of syllabus					

1997 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	4				
Photosynthesis	11				9
Respiration					
Human nutrition	20	8			
Human gaseous exchange	30	10		35	
Population dynamics					
Plant hormones					
Plant water relations	32		23		
Human excretion	30	4	16	4	
Human co-ordination	48		4		26
Thermo-regulation	22				
Human circulation	48	6			10
Osmoregulation					
Out of syllabus					

1997 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	4				
Photosynthesis	11				9
Respiration					
Human nutrition	50	18		35	
Human gaseous exchange					
Population dynamics					
Plant hormones					
Plant water relations	32		23		
Human excretion	30	4	16	4	
Human co-ordination	48		4		26
Thermo-regulation	22				
Human circulation	48	6			10
Osmoregulation					
Out of syllabus					

Figure 6.7 continued on next page

1997 Profile 3

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	4				
Photosynthesis	11				9
Respiration					
Human nutrition	20	8			
Human gaseous exchange	30	10		35	
Population dynamics					
Plant hormones					
Plant water relations	32		23		
Human excretion	30	4	16	4	
Human co-ordination	48		4		26
Thermo-regulation	9	13			
Human circulation	48	2			10
Osmoregulation					
Out of syllabus					

1997 Profile 4

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	4				
Photosynthesis	11				9
Respiration					
Human nutrition	50	18		35	
Human gaseous exchange					
Population dynamics					
Plant hormones					
Plant water relations	32		23		
Human excretion	30	4	16	4	
Human co-ordination	48		4		26
Thermo-regulation	9	13			
Human circulation	48	6			10
Osmoregulation					
Out of syllabus					

1997 Profile 5

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	4				
Photosynthesis	11				9
Respiration					
Human nutrition	20	8			
Human gaseous exchange	30	10		35	
Population dynamics					
Plant hormones					
Plant water relations	32		23		
Human excretion	28		26		
Human co-ordination	48		4		26
Thermo-regulation	22				
Human circulation	48	6			10
Osmoregulation					
Out of syllabus					

1997 Profile 6

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	4				
Photosynthesis	11				9
Respiration					
Human nutrition	50	18		35	
Human gaseous exchange					
Population dynamics					
Plant hormones					
Plant water relations	32		23		
Human excretion	28		26		
Human co-ordination	48		4		26
Thermo-regulation	22				
Human circulation	48	6			10
Osmoregulation					
Out of syllabus					

1997 Profile 7

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	4				
Photosynthesis	11				9
Respiration					
Human nutrition	20	8			
Human gaseous exchange	30	10		35	
Population dynamics					
Plant hormones					
Plant water relations	32		23		
Human excretion	28		26		
Human co-ordination	48		4		26
Thermo-regulation	9	13			
Human circulation	48	6			10
Osmoregulation					
Out of syllabus					

Figure 6.7 continued on next page

1997 Profile 8

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	4				
Photosynthesis	11				9
Respiration					
Human nutrition	50	18		35	
Human gaseous exchange					
Population dynamics					
Plant hormones					
Plant water relations	32		23		
Human excretion	28		26		
Human co-ordination	48		4		26
Thermo-regulation	9	13			
Human circulation	48	6			10
Osmoregulation					
Out of syllabus					

1998 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	24		8		4
Photosynthesis				9	
Respiration	9				
Human nutrition	30		2		11
Human gaseous exchange	25	6	18		6
Population dynamics					
Plant hormones					
Plant water relations	29	3	24		5
Human excretion	24	8		28	
Human co-ordination					
Thermo-regulation	9		10		
Human circulation	14		10		4
Osmoregulation					
Out of syllabus					

1998 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	24		8		4
Photosynthesis				9	
Respiration	9				
Human nutrition	30		2		11
Human gaseous exchange	25	6	18		6
Population dynamics					
Plant hormones					
Plant water relations	29	3	24		5
Human excretion					
Human co-ordination	24	8		28	
Thermo-regulation	9		10		
Human circulation	14		10		4
Osmoregulation					
Out of syllabus					

1998 Profile 3

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	24		8		4
Photosynthesis				9	
Respiration	9				
Human nutrition	30		2		11
Human gaseous exchange	25	6	18		6
Population dynamics					
Plant hormones					
Plant water relations	29	3	24		5
Human excretion					
Human co-ordination					
Thermo-regulation	9		10		
Human circulation	14		10		4
Osmoregulation					
Out of syllabus		8			52

1998 Profile 4

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	24		8		4
Photosynthesis				9	
Respiration	9				
Human nutrition	30		2		11
Human gaseous exchange	25	6	18		6
Population dynamics					
Plant hormones					
Plant water relations	39	3	14		5
Human excretion	24	8		28	
Human co-ordination					
Thermo-regulation	9		10		
Human circulation	14		10		4
Osmoregulation					
Out of syllabus					

Figure 6.7 continued on next page

1998 Profile 5

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	24		8		4
Photosynthesis				9	
Respiration	9				
Human nutrition	30		2		11
Human gaseous exchange	25	6	18		6
Population dynamics					
Plant hormones					
Plant water relations	39	3	14		5
Human excretion					
Human co-ordination	24	8		28	
Thermo-regulation	9		10		
Human circulation	14		10		4
Osmoregulation					
Out of syllabus					

1998 Profile 6

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	24		8		4
Photosynthesis				9	
Respiration	9				
Human nutrition	30		2		11
Human gaseous exchange	25	6	18		6
Population dynamics					
Plant hormones					
Plant water relations	39	3	14		5
Human excretion					
Human co-ordination					
Thermo-regulation	9		10		
Human circulation	14		10		4
Osmoregulation					
Out of syllabus		8			52

1999 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	13		15		12
Photosynthesis					
Respiration	12	4	2		6
Human nutrition	28				
Human gaseous exchange	5		29	6	
Population dynamics					
Plant hormones					
Plant water relations	24	8	28		
Human excretion	26	2		5	
Human co-ordination	29		22		
Thermo-regulation	2				
Human circulation	13		5	12	12
Osmoregulation					
Out of syllabus					

1999 Profile 2

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	13		15		12
Photosynthesis					
Respiration	12	4	2		6
Human nutrition	28				
Human gaseous exchange	5		29	6	
Population dynamics					
Plant hormones					
Plant water relations					
Human excretion	26	2		5	
Human co-ordination	29		22		
Thermo-regulation	26	8	28		
Human circulation	13		5	12	12
Osmoregulation					
Out of syllabus					

1999 Profile 3

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	13		7		20
Photosynthesis					
Respiration	12	4	2		6
Human nutrition	28				
Human gaseous exchange	5		29	6	
Population dynamics					
Plant hormones					
Plant water relations	24	8	28		
Human excretion	26	2		5	
Human co-ordination	29		22		
Thermo-regulation	2				
Human circulation	13		5	12	12
Osmoregulation					
Out of syllabus					

Figure 6.7 continued on next page

	1999 Profile 4					2000 Profile 1					2000 Profile 2					2000 Profile 3					2000 Profile 4				
	Memorize	Procedure	Explain	Analysis	Application	Memorize	Procedure	Explain	Analysis	Application	Memorize	Procedure	Explain	Analysis	Application	Memorize	Procedure	Explain	Analysis	Application	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	13		7		20	22	2	6	2		22	2	6	2		22	2	6	2		22	2	6	2	
Photosynthesis						16	4	18		4	4		4		4	16	4	18		4	4		4		4
Respiration	12	4	2		6	12	4	14								12	4	14							
Human nutrition	28					20		12			32	4	26			20		12			32	4	26		
Human gaseous exchange	5		29	6		4		20	4	12	16	4	34	4	12	4		20	4	12	16	4	34	4	12
Population dynamics																									
Plant hormones																									
Plant water relations						8	10			12	8	10			12	8	10			12	8	10			12
Human excretion	26	2		5		9		9	17		9		9	17		9		9	17		9		9	17	
Human co-ordination	29		22			34		12			34		12			34		12			34		12		
Thermo-regulation	24	8	28			10					10					10					10				
Human circulation	13		5	12	12	12		4	2		12		4	2		12		4	2		12		4	2	
Osmoregulation									1	4				1	4				1	4				1	4
Out of syllabus																									

Figure 6.7 continued on next page

2000 Profile 5

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	22	2	6	2	
Photosynthesis	16	4	18		4
Respiration	22	4	14		
Human nutrition	20		12		
Human gaseous exchange	4		10	4	12
Population dynamics					
Plant hormones					
Plant water relations	8	10			12
Human excretion	9		9	17	
Human co-ordination	34		12		
Thermo-regulation	10				
Human circulation	12		4	2	
Osmoregulation				1	4
Out of syllabus					

2000 Profile 6

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	22	2	6	2	
Photosynthesis	4		4		4
Respiration	10				
Human nutrition	32	4	26		
Human gaseous exchange	16	4	24	4	12
Population dynamics					
Plant hormones					
Plant water relations	8	10			12
Human excretion	9		9	17	
Human co-ordination	34		12		
Thermo-regulation	10				
Human circulation	12		4	2	
Osmoregulation			1	4	
Out of syllabus					

2000 Profile 7

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	22	2	6	2	
Photosynthesis	16	4	18		4
Respiration	22	4	14		
Human nutrition	20		12		
Human gaseous exchange	4		10	4	12
Population dynamics					
Plant hormones					
Plant water relations	8	10			12
Human excretion	9		9	17	
Human co-ordination	34		12		
Thermo-regulation	10				
Human circulation	12		4	2	
Osmoregulation			1	4	
Out of syllabus					

2000 Profile 8

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	22	2	6	2	
Photosynthesis	4		4		4
Respiration	10				
Human nutrition	32	4	26		
Human gaseous exchange	16	4	24	4	12
Population dynamics					
Plant hormones					
Plant water relations	8	10			12
Human excretion	9		9	17	
Human co-ordination	34		12		
Thermo-regulation	10				
Human circulation	12		4	2	
Osmoregulation				1	4
Out of syllabus					

2001 Profile 1

	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	14	6	4		
Photosynthesis					
Respiration	9		4	4	
Human nutrition	38	4	15	14	12
Human gaseous exchange	12	6	2	16	2
Population dynamics					
Plant hormones					
Plant water relations	7		20		8
Human excretion	4	6	13		12
Human co-ordination	32			6	
Thermo-regulation	4			6	2
Human circulation	14		4		
Osmoregulation					
Out of syllabus					

Figure 6.7 continued on next page

	2001 Profile 2					2002 Profile 1					2002 Profile 2					2003 Profile 1					2003 Profile 2				
	Memorize	**Procedure**	**Explain**	**Analysis**	**Application**	**Memorize**	**Procedure**	**Explain**	**Analysis**	**Application**	**Memorize**	**Procedure**	**Explain**	**Analysis**	**Application**	**Memorize**	**Procedure**	**Explain**	**Analysis**	**Application**	**Memorize**	**Procedure**	**Explain**	**Analysis**	**Application**
Biochemistry	14	6	4			16	4	20	8	2	16	4	20	8	2	16		2	5		16		2	5	
Photosynthesis						12					12										24	8	28		
Respiration	9		4	4		20			6		20			6		6		10	8		6		10	8	
Human nutrition	26		15		12	28	7	6	12	11	28	7	6	12	11	6	2	18	5	11	6	2	18	5	11
Human gaseous exchange		2	2	2	2	15	8	2			15	8	2			13	6	6			13	6	6		
Population dynamics																									
Plant hormones																									
Plant water relations	19	4	20	14	8	50	10	35			20					5		11	12		5		11	12	
Human excretion	16	10	13	14	12	26	2	6			26	2	6			11	4	4	3	14	11	4	4	3	14
Human co-ordination	32			6		22		4			52	10	35			36	4	14	3	4	24			3	4
Thermo-regulation	4			6	2	4		16	6	4	4		16	6	4	14	4	14			2				
Human circulation	14		4			18		12		8	18		12		8	19	2	6	2		19	2	6	2	
Osmoregulation																									
Out of syllabus																									

Figure 6.7 continued on next page

	2004 Profile 1					2004 Profile 2					2005 Profile 1					2005 Profile 2					2006 Profile 1				
	Memorize	Procedure	Explain	Analysis	Application	Memorize	Procedure	Explain	Analysis	Application	Memorize	Procedure	Explain	Analysis	Application	Memorize	Procedure	Explain	Analysis	Application	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	22		9		2	22		9		2	13		4		25	13		4		25	6			3	6
Photosynthesis	20	4	15		10	8		1		10	2		2		2	26	8	30		2	4				
Respiration	3					3					5					5					7		7		3
Human nutrition	12		6	2	4	12		6	2	4	9		19		6	9		19		6	35	8	38		14
Human gaseous exchange	15	4	23			3		9			13		4		7	13		4		7	12	4	2		
Population dynamics	12	4	6	7	5	12	4	6	7	5	10	4	8		6	10	4	8		6	4	6			12
Plant hormones																									
Plant water relations	2	4	6	3	14	2	4	6	3	14	4	6	5		11	4	6	5		11	8	2	6		8
Human excretion	7	2	3		23	7	2	3		23	35	8	28			11					9	2	11	2.5	
Human co-ordination	25		3			49	8	31			29	2	8		1	29	2	8		1	23	9	13		21
Thermo-regulation	11	2	8		2	11	2	8		2	5	2			17	5	2			17	12			2.5	
Human circulation																									
Osmoregulation																									
Out of syllabus																									

Figure 6.7 continued on next page

	2006 Profile 2					2007 Profile 1					2007 Profile 2				
	Memorize	Procedure	Explain	Analysis	Application	Memorize	Procedure	Explain	Analysis	Application	Memorize	Procedure	Explain	Analysis	Application
Biochemistry	6			3	6	13	3	6		8	13	3	6		8
Photosynthesis	4					9	3	5			9	3	5		
Respiration	7		7		3	14	5	3			14	5	3		
Human nutrition	11		10		14	16					16				
Human gaseous exchange	12	4	2				6	5		7	12	10	19		7
Population dynamics	28	14	28		12	6	7	2		23	6	7	2		23
Plant hormones															
Plant water relations	8	2	6		8	32	4	14			20				
Human excretion	9	2	11	2.5		12	4	14			12	4	14		
Human co-ordination	23	9	13		21	25	5	8		29	25	5	8		29
Thermo-regulation	12			2.5				10		2			10		2
Human circulation															
Osmoregulation															
Out of syllabus															

Figure 6.7 Variation in the composition of content (raw scores) 1994-2007 IEB Biology HG examinations. Shaded cells indicate differences between different profiles within a year.

a. ***Similarity of content***

Porter (2002) described the development of an alignment index which measured the similarity between the content emphases in distinct curricular areas based on the two-dimensional model of content (topic and cognitive demand) described above. He advocated that the alignment index be used together with content descriptions from test papers to answer questions such as, for example, "[w]hen alignment is low, what is it about content that yields the low alignment? And what are the areas in which alignment exists?" (Porter, 2002, p.7). While alignment is a concept that originated in the USA because of the legal requirements of NCLB (Chapter 3, Section 3.4.4), similar questions must be asked in this study about comparisons of content between different SC examinations, that is, what is it about content that makes some papers more similar than other papers? Porter's similarity index was re-conceptualized as the similarity index (SI) to be used in this study to compare the content emphases between question papers (Chapter 5, Section 5.2.3.2).

Similarity values for pairs of the content distributions discussed above and shown in Figures 6.6 and 6.7 were calculated using the SI. Within each year, the greatest similarity in content emphases were consistently between the profiles within each examination body when multiple profiles existed, for both HG and SG question papers (Tables 6.22 - 6.27). For example, despite the dissimilarities in the content emphasized by CED 1994 HG Profile 1 and Profile 2 described above (Figure 6.6), these profiles are 85% similar (Table 6.22), whereas the closest either profile is to any other CED-WCED-DoE examination is 65% (CED 1994 Profile 1 and WCED 1996 HG Profile 2) (Table 6.22). This suggests that choices of questions within an examination constitute less of a challenge to equivalence if similarity when content is considered necessary for equivalence. The DoE policy introduced in 2001 which required DoE examinations to examine all topics in given proportions, succeeded in minimizing differences in SI values between DoE HG and SG examination question papers, both within and between years, from 2001 to 2007 (Tables 6.22 & 6.23). Comparisons were made of corresponding IEB HG question papers (Tables 6.24 & 6.25) showed highly variable SI values. Comparisons of content emphases between HG and SG question papers showed a greater similarity between the DoE HG and SG question papers than was observed between the IEB HG and SG question papers, both within and between years (Tables 6.26 & 6.27). These differences and ranges of SI values summarized in Table 6.28 were expected, given that post-2000 the DoE had a policy that required that various topics be examined in the same proportions for both HG and SG whereas the IEB did not have such policies. Given that differentiation of the curriculum of subjects into HG and SG was intended to make teaching and learning accessible to two groups of students, each with different sets of strengths and weaknesses as well as different career aspirations, the wisdom of the DoE policy of

Table 6.22 Summary of the content similarity, measured by the similarity index (SI), 1994, 1995 CED,1996-2000 WCED and 2001-2007 national DoE Biology HG question papers.

		Year																				
		1994		1995		1996		1997		1998		1999		2000		2001	2002	2003	2004	2005	2006	2007
Year	**Profile**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
1994	**1**	1.00	0.85	0.55	0.60	0.58	0.65	0.55	0.52	0.56	0.50	0.55	0.58	0.55	0.50	0.53	0.54	0.53	0.47	0.49	0.51	0.49
	2		1.00	0.57	0.63	0.61	0.58	0.59	0.56	0.64	0.54	0.56	0.64	0.56	0.49	0.54	0.55	0.54	0.48	0.50	0.49	0.47
1995	**1**			1.00	0.85	0.61	0.64	0.53	0.52	0.60	0.52	0.56	0.58	0.60	0.67	0.59	0.58	0.61	0.55	0.61	0.59	0.53
	2				1.00	0.66	0.60	0.55	0.57	0.64	0.56	0.65	0.68	0.64	0.59	0.62	0.57	0.60	0.54	0.62	0.61	0.60
1996	**1**					1.00	0.84	0.63	0.69	0.65	0.57	0.65	0.61	0.60	0.57	0.61	0.53	0.53	0.49	0.53	0.51	0.55
	2						1.00	0.61	0.63	0.57	0.49	0.57	0.57	0.56	0.62	0.55	0.53	0.51	0.44	0.48	0.48	0.50
1997	**1**							1.00	0.85	0.67	0.62	0.56	0.62	0.62	0.56	0.55	0.52	0.55	0.46	0.48	0.41	0.48
	2								1.00	0.63	0.58	0.60	0.55	0.53	0.58	0.54	0.48	0.49	0.45	0.50	0.41	0.51
1998	**1**									1.00	0.85	0.61	0.68	0.61	0.57	0.58	0.56	0.54	0.55	0.55	0.51	0.57
	2										1.00	0.56	0.60	0.55	0.50	0.55	0.50	0.51	0.48	0.49	0.45	0.52
1999	**1**											1.00	0.87	0.65	0.55	0.64	0.56	0.59	0.54	0.57	0.56	0.62
	2												1.00	0.65	0.54	0.58	0.58	0.58	0.54	0.53	0.51	0.55
2000	**1**													1.00	0.85	0.59	0.55	0.63	0.55	0.57	0.55	0.54
	2														1.00	0.50	0.50	0.57	0.50	0.54	0.47	0.46
2001	**1**															1.00	0.60	0.69	0.67	0.61	0.68	0.71
2002	**1**																1.00	0.65	0.69	0.62	0.64	0.65
2003	**1**																	1.00	0.66	0.66	0.66	0.69
2004	**1**																		1.00	0.62	0.64	0.64
2005	**1**																			1.00	0.67	0.66
2006	**1**																				1.00	0.69
2007	**1**																					1.00

Table 6.23 Summary of the content similarity, measured by the similarity index (SI), 2001-2007 national DoE Biology SG question papers.

Year	2001	2002	2003	2004	2005	2006	2007
2001	1.00	0.73	0.71	0.75	0.71	0.73	0.69
2002		1.00	0.70	0.72	0.69	0.78	0.69
2003			1.00	0.75	0.69	0.78	0.69
2004				1.00	0.74	0.81	0.70
2005					1.00	0.69	0.74
2006						1.00	0.72
2007							1.00

Table 6.24 Summary of the content similarity, measured by the similarity index (SI), 1994-2007 IEB Biology HG question papers.

		Year																					
		1994		1995		1996				1997								1998					
Year	**Profile**	**1**	**2**	**1**	**2**	**1**	**2**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**
1994	**1**	1.00	0.85	0.46	0.50	0.40	0.32	0.40	0.33	0.24	0.25	0.25	0.26	0.24	0.25	0.25	0.26	0.34	0.39	0.34	0.34	0.38	0.33
	2		1.00	0.47	0.51	0.38	0.31	0.38	0.32	0.29	0.28	0.30	0.29	0.31	0.29	0.32	0.30	0.36	0.32	0.27	0.35	0.31	0.27
1995	**1**			1.00	0.85	0.46	0.41	0.44	0.41	0.32	0.30	0.32	0.30	0.32	0.30	0.32	0.30	0.34	0.42	0.32	0.34	0.42	0.32
	2				1.00	0.40	0.35	0.40	0.38	0.36	0.34	0.36	0.34	0.36	0.34	0.36	0.34	0.34	0.43	0.32	0.34	0.43	0.32
1996	**1**					1.00	0.81	0.97	0.81	0.44	0.39	0.44	0.39	0.43	0.38	0.43	0.38	0.49	0.51	0.44	0.49	0.51	0.44
	2						1.00	0.78	0.97	0.39	0.33	0.38	0.33	0.38	0.32	0.37	0.32	0.46	0.48	0.41	0.46	0.48	0.41
	5							1.00	0.81	0.44	0.39	0.44	0.39	0.43	0.38	0.43	0.38	0.49	0.51	0.44	0.49	0.51	0.44
	6								1.00	0.42	0.36	0.41	0.35	0.41	0.35	0.40	0.34	0.46	0.48	0.41	0.46	0.48	0.41
1997	**1**									1.00	0.81	0.97	0.78	0.98	0.79	0.94	0.76	0.47	0.45	0.38	0.46	0.44	0.36
	2										1.00	0.78	0.97	0.79	0.98	0.76	0.94	0.42	0.40	0.33	0.41	0.39	0.31
	3											1.00	0.81	0.94	0.76	0.98	0.79	0.47	0.45	0.37	0.45	0.43	0.36
	4												1.00	0.76	0.94	0.79	0.98	0.42	0.40	0.32	0.40	0.38	0.31
	5													1.00	0.81	0.97	0.78	0.45	0.45	0.38	0.43	0.44	0.36
	6														1.00	0.78	0.97	0.40	0.40	0.33	0.38	0.39	0.31
	7															1.00	0.81	0.44	0.45	0.37	0.43	0.43	0.36
	8																1.00	0.39	0.40	0.32	0.38	0.38	0.31
1998	**1**																	1.00	0.81	0.81	0.97	0.78	0.78
	2																		1.00	0.81	0.78	0.97	0.78
	3																			1.00	0.78	0.78	0.97
	4																				1.00	0.81	0.81
	5																					1.00	0.81
	6																						1.00

Table 6.24 continued

		Year																					
		1999				2000				2001		2002		2003		2004		2005		2006		2007	
Year	**Profile**	**1**	**2**	**3**	**4**	**1, 3**	**2, 4**	**5, 7**	**6, 8**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
1994	**1**	0.43	0.39	0.43	0.39	0.45	0.42	0.45	0.44	0.39	0.36	0.34	0.38	0.37	0.36	0.42	0.46	0.47	0.52	0.46	0.38	0.47	0.43
	2	0.44	0.40	0.44	0.40	0.41	0.33	0.41	0.35	0.33	0.37	0.33	0.37	0.35	0.34	0.40	0.45	0.54	0.47	0.38	0.39	0.52	0.48
1995	**1**	0.40	0.37	0.34	0.34	0.47	0.44	0.47	0.47	0.42	0.43	0.40	0.40	0.36	0.34	0.59	0.59	0.47	0.45	0.43	0.48	0.56	0.45
	2	0.45	0.42	0.40	0.51	0.51	0.48	0.50	0.51	0.42	0.43	0.38	0.56	0.43	0.34	0.59	0.73	0.50	0.48	0.47	0.52	0.56	0.59
1996	**1**	0.47	0.38	0.50	0.40	0.48	0.46	0.45	0.46	0.45	0.44	0.46	0.41	0.45	0.55	0.53	0.38	0.38	0.43	0.36	0.34	0.45	0.48
	2	0.39	0.57	0.41	0.59	0.38	0.42	0.35	0.42	0.39	0.36	0.38	0.39	0.51	0.41	0.47	0.39	0.35	0.32	0.33	0.31	0.33	0.42
	5	0.47	0.38	0.50	0.40	0.49	0.46	0.46	0.46	0.45	0.44	0.46	0.41	0.45	0.58	0.56	0.38	0.38	0.47	0.36	0.34	0.45	0.48
	6	0.39	0.52	0.41	0.59	0.41	0.44	0.38	0.44	0.39	0.36	0.41	0.42	0.51	0.44	0.50	0.41	0.36	0.35	0.35	0.32	0.36	0.45
1997	**1**	0.49	0.41	0.49	0.41	0.40	0.45	0.40	0.45	0.50	0.47	0.53	0.50	0.48	0.42	0.36	0.34	0.38	0.31	0.47	0.43	0.48	0.47
	2	0.49	0.41	0.49	0.41	0.37	0.42	0.39	0.42	0.51	0.50	0.52	0.50	0.44	0.37	0.30	0.34	0.34	0.27	0.49	0.38	0.47	0.41
	3	0.49	0.40	0.49	0.40	0.39	0.44	0.39	0.44	0.50	0.47	0.53	0.50	0.47	0.42	0.35	0.34	0.39	0.32	0.45	0.42	0.48	0.47
	4	0.49	0.40	0.49	0.40	0.38	0.42	0.38	0.42	0.51	0.50	0.52	0.50	0.43	0.37	0.30	0.33	0.34	0.27	0.47	0.36	0.41	0.41
	5	0.47	0.39	0.47	0.39	0.39	0.44	0.39	0.44	0.47	0.46	0.52	0.50	0.46	0.40	0.35	0.34	0.39	0.31	0.45	0.42	0.48	0.46
	6	0.47	0.39	0.47	0.39	0.38	0.41	0.38	0.41	0.50	0.48	0.51	0.49	0.42	0.35	0.30	0.33	0.35	0.27	0.47	0.37	0.46	0.40
	7	0.47	0.38	0.47	0.38	0.38	0.43	0.38	0.43	0.49	0.46	0.52	0.50	0.45	0.40	0.35	0.33	0.40	0.32	0.43	0.40	0.48	0.46
	8	0.47	0.38	0.47	0.38	0.37	0.41	0.37	0.41	0.50	0.48	0.51	0.49	0.41	0.35	0.29	0.33	0.35	0.27	0.45	0.35	0.46	0.40
1998	**1**	0.60	0.49	0.59	0.49	0.48	0.53	0.46	0.56	0.50	0.59	0.64	0.51	0.46	0.41	0.43	0.35	0.39	0.33	0.39	0.34	0.44	0.44
	2	0.58	0.47	0.57	0.47	0.48	0.53	0.45	0.56	0.57	0.56	0.62	0.54	0.50	0.43	0.48	0.42	0.37	0.37	0.45	0.39	0.48	0.48
	3	0.50	0.39	0.50	0.39	0.40	0.45	0.38	0.48	0.47	0.46	0.57	0.44	0.40	0.35	0.40	0.32	0.29	0.29	0.35	0.29	0.39	0.39
	4	0.57	0.49	0.59	0.49	0.48	0.53	0.46	0.56	0.48	0.56	0.64	0.51	0.46	0.41	0.43	0.35	0.39	0.33	0.39	0.34	0.46	0.44
	5	0.54	0.47	0.54	0.47	0.48	0.53	0.45	0.56	0.54	0.53	0.62	0.54	0.50	0.43	0.48	0.42	0.37	0.37	0.45	0.39	0.49	0.48
	6	0.47	0.37	0.47	0.39	0.40	0.45	0.38	0.48	0.45	0.44	0.57	0.44	0.40	0.35	0.40	0.32	0.29	0.29	0.35	0.29	0.40	0.39

Table 6.24 continued

		Year																					
		1999				**2000**				**2001**		**2002**		**2003**		**2004**		**2005**		**2006**		**2007**	
Year	**Profile**	**1**	**2**	**3**	**4**	**1, 3**	**2, 4**	**5, 7**	**6, 8**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
1999	**1**	1.00	0.81	0.98	0.79	0.55	0.55	0.52	0.57	0.48	0.56	0.63	0.58	0.44	0.38	0.39	0.40	0.43	0.38	0.41	0.36	0.50	0.50
	2		1.00	0.79	0.98	0.53	0.53	0.49	0.54	0.40	0.42	0.48	0.58	0.49	0.33	0.41	0.43	0.40	0.34	0.39	0.34	0.40	0.47
	3			1.00	0.81	0.55	0.55	0.52	0.57	0.48	0.56	0.63	0.58	0.44	0.38	0.39	0.40	0.43	0.38	0.41	0.36	0.50	0.50
	4				1.00	0.53	0.53	0.49	0.54	0.40	0.42	0.46	0.55	0.49	0.33	0.40	0.42	0.42	0.37	0.39	0.34	0.40	0.47
2000	**1, 3**					1.00	0.81	0.97	0.84	0.51	0.57	0.45	0.51	0.51	0.54	0.57	0.51	0.48	0.55	0.47	0.44	0.48	0.53
	2, 4						1.00	0.78	0.97	0.57	0.57	0.45	0.51	0.53	0.46	0.54	0.48	0.52	0.50	0.56	0.44	0.40	0.48
	5, 7							1.00	0.81	0.51	0.57	0.46	0.52	0.51	0.54	0.54	0.51	0.48	0.55	0.47	0.44	0.49	0.51
	6, 8								1.00	0.60	0.60	0.49	0.54	0.55	0.48	0.55	0.49	0.53	0.52	0.58	0.46	0.44	0.51
2001	**1**									1.00	0.80	0.59	0.58	0.62	0.59	0.44	0.42	0.48	0.41	0.56	0.45	0.42	0.42
	2										1.00	0.59	0.55	0.62	0.58	0.42	0.43	0.50	0.41	0.51	0.44	0.49	0.43
2002	**1**											1.00	0.81	0.54	0.51	0.43	0.39	0.41	0.38	0.44	0.39	0.56	0.50
	2												1.00	0.61	0.50	0.43	0.54	0.44	0.41	0.50	0.45	0.51	0.54
2003	**1**													1.00	0.80	0.49	0.53	0.47	0.45	0.54	0.51	0.42	0.43
	2														1.00	0.55	0.40	0.42	0.58	0.46	0.43	0.40	0.41
2004	**1**															1.00	0.80	0.53	0.63	0.45	0.49	0.47	0.52
	2																1.00	0.53	0.54	0.47	0.51	0.46	0.45
2005	**1**																	1.00	0.80	0.54	0.55	0.51	0.52
	2																		1.00	0.50	0.52	0.49	0.50
2006	**1**																			1.00	0.80	0.55	0.56
	2																				1.00	0.55	0.56
2007	**1**																					1.00	0.90
	2																						1.00

Table 6.25 Summary of the content similarity, measured by the similarity index (SI), 2001-2007 IEB Biology SG question papers.

		Year													
		2001		2002		2003		2004		2005		2006		2007	
Year	**Profile**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
2001	**1**	1.00	0.91	0.56	0.54	0.49	0.44	0.45	0.43	0.44	0.50	0.48	0.49	0.50	0.54
	2		1.00	0.58	0.56	0.51	0.45	0.45	0.45	0.51	0.56	0.53	0.54	0.56	0.60
2002	**1**			1.00	0.92	0.54	0.51	0.53	0.53	0.52	0.57	0.47	0.48	0.53	0.47
	2				1.00	0.53	0.50	0.52	0.52	0.52	0.51	0.50	0.51	0.51	0.49
2003	**1**					1.00	0.91	0.55	0.62	0.48	0.50	0.43	0.51	0.60	0.53
	2						1.00	0.53	0.60	0.54	0.47	0.47	0.53	0.54	0.47
2004	**1**							1.00	0.91	0.61	0.61	0.56	0.59	0.62	0.56
	2								1.00	0.59	0.59	0.50	0.54	0.63	0.57
2005	**1**									1.00	0.91	0.58	0.62	0.63	0.57
	2										1.00	0.52	0.56	0.64	0.58
2006	**1**											1.00	0.91	0.57	0.60
	2												1.00	0.61	0.64
2007	**1**													1.00	0.91
	2														1.00

Table 6.26 Summary of the content range similarity, measured by the similarity index (SI), national DoE, 2001 to 2007 Biology HG and SG question papers.

		SG						
		Year						
	Year	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
HG	**2001**	0.62	0.73	0.62	0.66	0.72	0.64	0.67
	2002	0.63	0.63	0.61	0.64	0.67	0.65	0.61
	2003	0.65	0.66	0.69	0.69	0.65	0.66	0.72
	2004	0.56	0.55	0.55	0.63	0.65	0.64	0.60
	2005	0.58	0.55	0.59	0.63	0.64	0.60	0.61
	2006	0.55	0.54	0.57	0.65	0.61	0.60	0.62
	2007	0.59	0.62	0.61	0.69	0.71	0.68	0.71

Table 6.27 Summary of the content range similarity, measured by the similarity index (SI), 2001-2007 IEB Biology HG and SG question papers.

			SG													
			Year													
			2001		2002		2003		2004		2005		2006		2007	
	Year	Profile	1	2	1	2	1	2	1	2	1	2	1	2	1	2
HG	2001	1	0.52	0.51	0.52	0.48	0.40	0.34	0.41	0.36	0.39	0.44	0.43	0.44	0.48	0.45
		2	0.54	0.52	0.51	0.48	0.43	0.42	0.42	0.38	0.45	0.45	0.47	0.48	0.48	0.47
	2002	1	0.58	0.58	0.52	0.51	0.51	0.45	0.40	0.40	0.41	0.44	0.45	0.47	0.52	0.54
		2	0.52	0.52	0.59	0.58	0.55	0.54	0.53	0.53	0.44	0.47	0.49	0.51	0.54	0.49
	2003	1	0.54	0.51	0.54	0.58	0.53	0.48	0.45	0.46	0.38	0.37	0.43	0.44	0.44	0.41
		2	0.51	0.57	0.54	0.52	0.43	0.36	0.33	0.38	0.36	0.43	0.34	0.35	0.42	0.40
	2004	1	0.50	0.57	0.53	0.52	0.43	0.35	0.47	0.51	0.52	0.54	0.49	0.51	0.50	0.46
		2	0.43	0.45	0.51	0.50	0.43	0.48	0.56	0.58	0.54	0.51	0.56	0.58	0.45	0.41
	2005	1	0.40	0.40	0.43	0.42	0.44	0.39	0.54	0.54	0.44	0.48	0.47	0.54	0.51	0.51
		2	0.42	0.50	0.45	0.44	0.35	0.30	0.47	0.50	0.46	0.50	0.49	0.47	0.50	0.50
	2006	1	0.48	0.49	0.37	0.39	0.37	0.34	0.48	0.44	0.40	0.42	0.53	0.54	0.53	0.54
		2	0.40	0.41	0.34	0.36	0.37	0.34	0.48	0.50	0.40	0.42	0.45	0.46	0.46	0.48
	2007	1	0.51	0.53	0.43	0.42	0.45	0.44	0.40	0.46	0.46	0.48	0.49	0.52	0.57	0.58
		2	0.50	0.53	0.51	0.49	0.47	0.41	0.42	0.48	0.43	0.49	0.44	0.47	0.60	0.55

Table 6.28 Summary of the content similarity, measured by the similarity index (SI), of Biology HG and SG question papers.

Question papers compared	Range of SI values between profiles of each question paper	Range of SI values between all profiles of different question papers
HG		
1994 All	0.81 to 0.91	0.30 to 0.70
1995 All	0.81 to 0.94	0.33 to 0.67
1996 All	0.78 to 0.93	0.27 to 0.70
1997 All	0.76 to 0.95	0.32 to 0.73
1998 All	0.78 to 0.97	0.32 to 0.68
1999 All	0.81 to 0.93	0.30 to 0.70
2000 All	0.78 to 0.97	0.35 to 0.70
2001 to 2007		
All	0.80 to 0.90	0.38 to 0.71
National DoE	–	0.60 to 0.71
IEB	0.80 to 0.90	0.38 to 0.59
1994 to 2007		
IEB	0.76 to 0.98	0.27 to 0.73
SG		
2001 to 2007		
All	0.91 to 0.92	0.43 to 0.75
National DoE	–	0.66 to 0.75
IEB	0.91 to 0.92	0.43 to 0.63
HG vs SG		
2001 to 2007		
National DoE	n/a	0.54 to 0.73
IEB	0.43 to 0.91	0.30 to 0.60

Note:
1. A ɓô means no unique profiles within each of the questions papers in the comparison.
2. Where a question paper had a choice of questions and each choice resulted in a different content profile, the profiles for that question paper were compared to each other.
3. All profiles of one question paper are compared to all profiles of other question papers in the comparison.
4. 2000 KZN Education Department ï combination of Paper 1 and Paper 2.

the same topic weighting in HG and SG question papers is questionable, given the different needs of students. If the policy was intended to bring consistency of emphases between the examinations of different years, for equivalence purposes, it failed because topics were examined at different levels of cognitive demand, which affects student performance in different ways (Sections 6.2.2.1 and 6.2.2.2).

Similarity between question papers (as indicated by SI values) on their own tell only part of the story about the various emphases in content that were observed in SC Biology examinations within and between years. The examples of similarity and differences between the SC Biology examinations discussed in this section illustrate the importance of using the content maps together with SI values for comparisons because they tell us different things about content.[217] Because the SI compares the proportions of marks between question papers, the different total number marks papers may carry could be overlooked, but tandem use of the content maps incorporates and compensates for the total number of marks per content category. Not converting content maps into topological maps as recommended by Porter (2002) keeps each cell distinct from neighbouring cells and the use of shading in the content maps helps to easily locate differences in content between corresponding profiles generated when an examination offers choices of questions which address particular content. Using this methodology, the content standards of examinations can be visually, and easily, compared within and between years as shown in Figures 6.6 and 6.7.

b. ***Breadth and depth of knowledge*** (Appendices 6.47 - 6.55)

Breadth of knowledge and depth of knowledge are two "dimensions of subject matter knowledge [which] figure centrally in any conception of mature disciplinary understanding" (Wineburg, 1997, p. 259). Schwartz et al. (2009, p. 799) described the emphasis on a wide range of topics, or an increased quantity of information, as "breadth" of knowledge, and levels of understanding (i.e., cognitive demand) as "depth" of knowledge. While both breadth and depth are necessary in curricula (Hirsch, 2001), the optimal balance between the relative degree of breadth and depth in science curricula has been debated for a long time (Anderson, 1995) and continues to be debated (Schwartz et al. 2009). While educators debate about whether students should be taught or assessed on a larger or smaller body of material, and the depth at which that material should be taught or assessed, there is no clear answer (Gandal, 1994). The explosion of scientific knowledge that took place in the 20th century means that there are many new potential topics and hence previous arguments for a broad curriculum are no longer admissible, given limited teaching time, nor necessarily desirable (Schwartz et al., 2009). Consequently, a number of recent books which have greatly influenced science curricula in the USA, have advocated the teaching of fewer topics in greater depth (AAAS, 1989, 1993; NRC 1999, 2007).

In a study of high-school leaving Biology examinations from six countries Gandal (1994) observed that while some countries emphasized breadth more than depth, other countries had

[217] The Porter index of similarity and the SI, have been criticized as being statistically "very simplistic" (Tim Dunne, personal communication May 8, 2011). It is for this very reason that I have chosen to use this index, because it conveys similarity in a way that could it can be easily understood by all involved with the SC examining processes. A second reason for using this index is that I could find no other indices in the literature which quantitatively compared content standards between tests.

the opposite approach, and some emphasized both hence requiring very long examinations. Gandal (1994) used France an example of a country where although the Biology curriculum taught was wide (many topics covered), the curriculum examined was narrow (few topics covered) and deep, illustrating that understanding examinations needs to happen with consideration of what is taught. French students had to learn all the topics for their school-leaving examinations as they did not know which topics would be examined (Gandal, 1994). Gandal argued that the French examinations were both rigorous and difficult, as they sacrificed neither breadth or depth. In South Africa, prior to 2001 students did not know which topics would be examined in the SC Biology examinations. From 2001 onwards students who wrote the national DoE examinations, knew both the topics and the proportions in which they would appear in each question paper. The IEB examinations had no policy with respect to breadth of knowledge. It could be argued that because the DoE stipulated, post-2000, the BOK of the SC Biology questions papers the papers became more predictable than comparable examinations of the IEB where any of the prescribed topics could be examined in any proportion. The DoE policy with a fixed BOK might also have resulted in teachers in DoE administered schools giving less attention to the teaching of topics with lower weightings in the examinations. While no rationale was provided by which to understand the DoE BOK weightings, it is not necessarily true that topics with lower weightings in the SC examinations were topics of least importance for high-school-leaving students (See Section 6.1.2.1).

Schmidt et al. (2005) compared the science curricula of almost 50 countries and found that top-achieving countries, defined in terms of student performance in TIMSS, covered far fewer topics with more coherence and depth, than did the USA. While breadth of coverage might result in higher test scores, if that is what the test emphasized (rather than depth of knowledge), studying at least one science topic in depth over a period of time in high school resulted in better student performances in introductory post-secondary courses (Schwartz et al., 2009).
Section 6.1.2.1 above described the Biology subject knowledge that was deemed *post hoc* to be worth knowing (i.e., the topics) by virtue of the fact that it had been examined in the SC Biology examinations analyzed in this study. Section 6.1.2.2 described how the students were expected to show evidence of how they could process this knowledge (i.e., performance expectations). Topics and performance expectations together conveyed the content standards (Section 6.1.2.3). The relationship between the knowledge conveyed by the topics (i.e., facts, concepts, processes, procedures) and how that knowledge was used (i.e., levels of cognitive demand) and the emphases on the distinct categories of these variables varied considerably between the different SC examination question papers, as is discussed below.

In order to explore the relationship between combinations of the breadth of knowledge (BOK) and the depth of knowledge (DOK), each of these variables was quantified for each question paper. Analyses of relationships between the BOK and DOK of the examinations set by different examining bodies for each of the years between 1994 and 2000 (Appendices 6.47 - 6.53) and for the DoE and the IEB for the period 2001 to 2007 (Appendices 6.54 - 6.55), showed no consistent discernable relationship between BOK and DOK. That is, there is no consistent evidence in this study that when fewer topics were examined the topics were examined, they were examined more deeply (Figure 6.8). The CED-WCED-DoE 1994 to 2000 HG examinations, with the exception of one of the profiles (i.e., WCED 1998 HG Profile 2), asked questions on fewer topics than did the post-2000 examinations. The WCED 1998 HG examination offered a choice of questions, one of which was an essay based on one topic only (Profile 2), and the other consisted of a series of shorter questions about three different topics (Appendix 6.33). Post-2000, the DoE 2005 HG examination was different in that it had many more Analyze questions than were found in the 2001 to 2004, and the 2006 and 2007 examinations, hence its higher DOK value (Figure 6.8 [A]). The IEB 1994 to 2007 HG examinations, except for some of the 2000 profiles (the content standards and structural aspects of question paper), covered a similar range of topics with similar emphases of cognitive demand (Figure 6.8 [B]), but generally fewer topics were examined than in the CED-WCED-DoE examinations (Figure 6.8 [A]). Four of the eight IEB 2000 HG profiles (Profiles 1, 3, 5 and 7) showed much higher BOK values than the other profiles in the same examination (Appendix 6.53) because of the greater choice of questions offered. Profiles 1, 3, 5 and 7 resulted in more topics being examined in similar proportions. Profiles 6 and 8 had the same number of topics but some were more heavily emphasized than other topics. Profiles 2 and 4 had one fewer topic *and* some topics more heavily emphasized than other topics. The various profiles of the IEB 2000 HG examinations illustrate the capacity of BOK to capture and reflect small changes in the spread and emphases of topics. The trend towards larger BOK observed in the CED-WCED-DoE question papers post-2000 is not surprising given that in 2001 DoE-set question papers were required to examine all topics specified in the CBS (Chapter 2, Table 2.8) (Table 6.17) resulting in a consistent BOK of approximately 9.38 (Figure 6.8) whereas IEB examiners could set questions on, and emphasize any topics (Figure 6.8; Table 6.17).

Generally, the DOE 2001-2007 SG examinations showed more breadth of content (Figure 6.8 [C]) than was found in the IEB examinations of the same years (Figure 6.8 [D]). This outcome was expected because of the DoE policy which specified the breadth of knowledge to be tested (Chapter 2, Table 2.8; Table 6.17). Both the DoE and the IEB 2001-2007 SG examinations showed lower DOK than the corresponding HG examinations in the same years, as was required by the CBS policy (Chapter 2, Table 2.8).

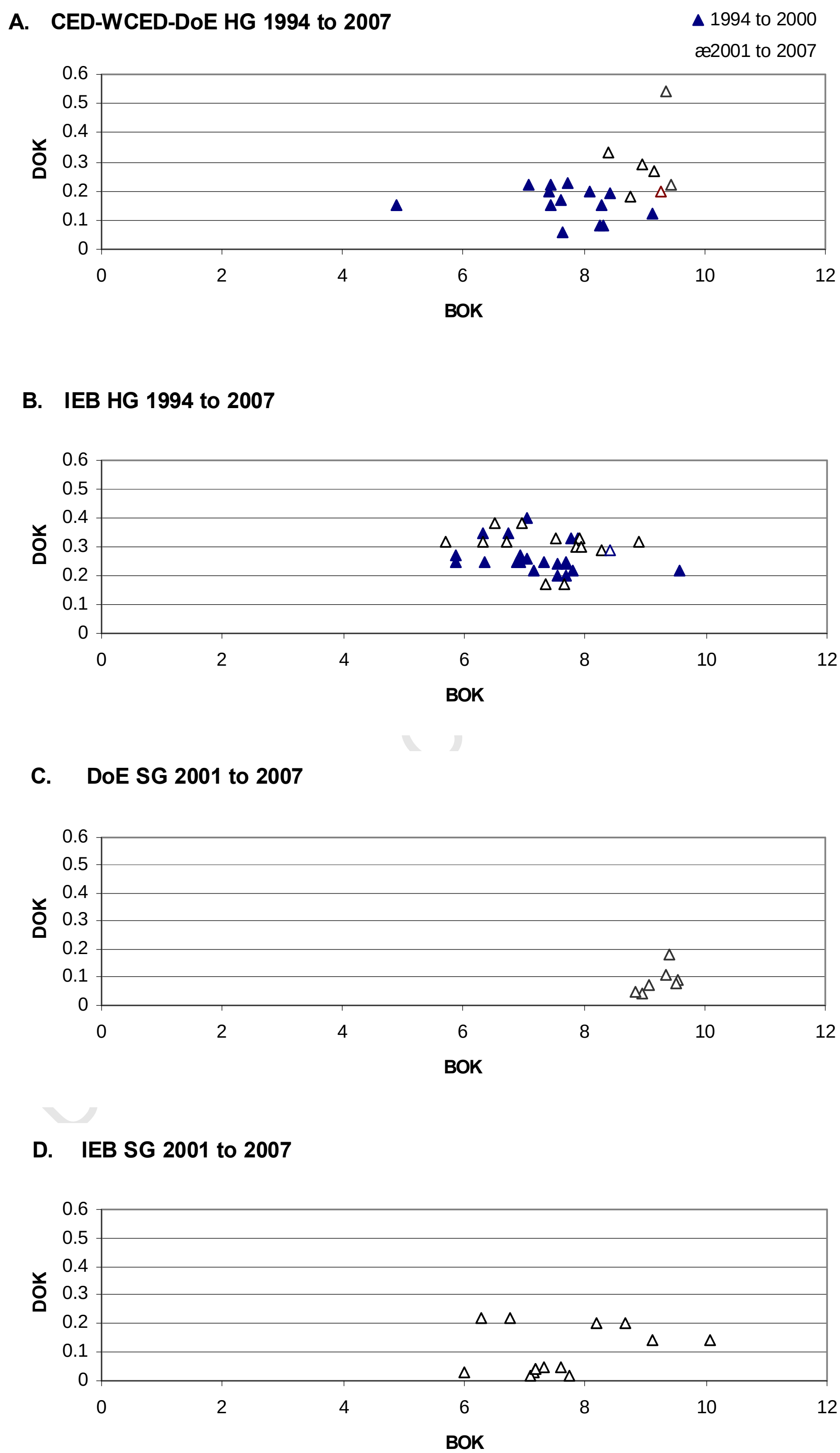


Figure 6.8 Relationships between breadth of knowledge (BOK) and depth of knowledge (DOK), 1994 to 2007.

The relationship between BOK and DOK and the effects, or lack thereof, of each on student performance in the 2005 and 2006 DoE examinations will be further explored and discussed in Section 6.2.2.1 and Section 6.2.2.2.

6.1.3 Comparisons of profiles (Appendix 6.56)

Section 6.1.1 described the structural features of the SC examination question papers, that is, the macrostructure and the structures of the questions and answers required by each question paper. Section 6.1.2 described, for each of the question papers, its content standards as combinations of topics and performance expectations. Both of these sections described considerable variation in these two characteristics when used individually to describe all the question papers analyzed. The current section examines whether or not a combination of the structural features and the content standard as the profile(s) of a question paper bring more understanding to comparisons of question papers.

The choice of question(s) within a question paper often resulted in different structural components and/or content standards depicted as multiple profiles for one question paper. Where a HG or SG examination consisted of more than one question paper (e.g., KZN HG 2000); national DoE HG &SG 2001 to 2007) the question papers were combined to form one profile. The resulting profiles were then subjected to exploratory cluster analyses in an attempt to capture the multivariate nature of examination question papers. Cluster analysis simultaneously considers the individual attributes (James & McCulloch, 1990). which comprise each profile and is therefore used in this study to give a visual overview of whether or not patterns of profiles of SC examinations that were not discernable when the attributes were examined individually, can be elicited over the fourteen years of this study.

Cluster analyses of the CED-WCED-DoE HG question papers (Figure 6.9), the IEB HG question papers (Figure 6.10), and the IEB and the DoE HG and SG question papers (Figure 6.11) showed that profiles for one examination session (year) were always more similar to each other than they were to other question papers in the comparison. While the pattern of similarity between profiles within the same examination was not surprising, given that the similarity between content was always greatest between such profiles (Section 6.1.2.3 a), different profiles of a single examination were often different in terms of their structural features (Section 6.1.1).

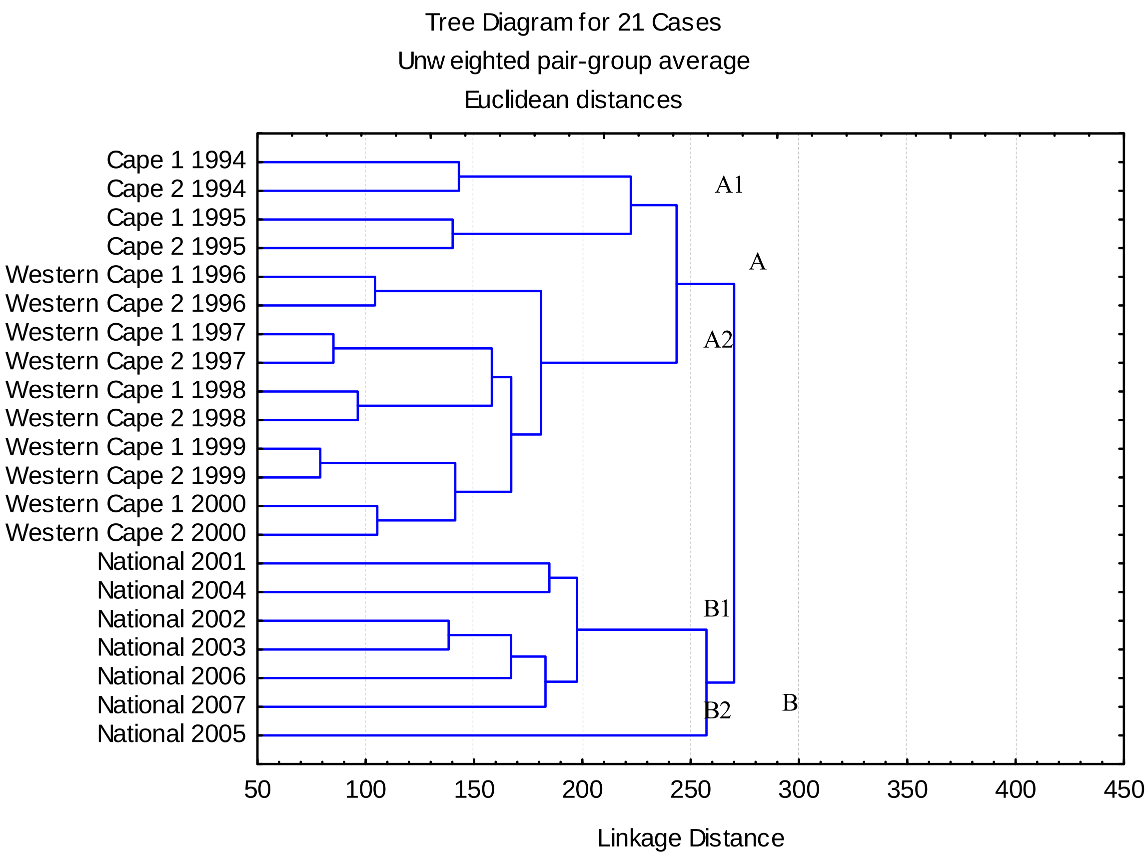


Figure 6.9 Results of a cluster analysis of the profiles of 1994, 1995 CED, 1996-2000 WCED and 2001-2007 national DoE SC Biology HG question papers.

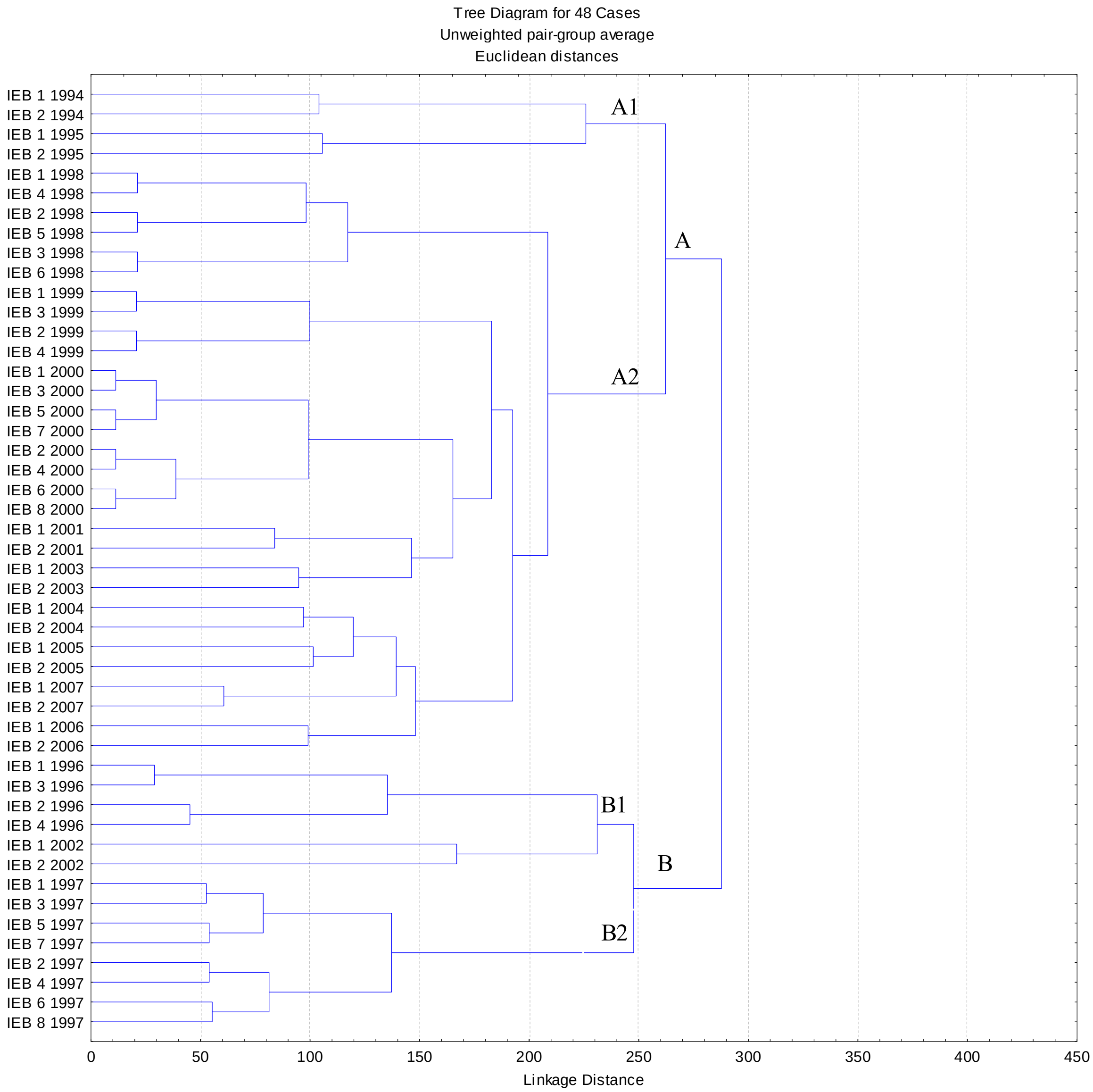


Figure 6.10 Results of a cluster analysis of the profiles of 1994-2007 IEB SC Biology HG question papers.

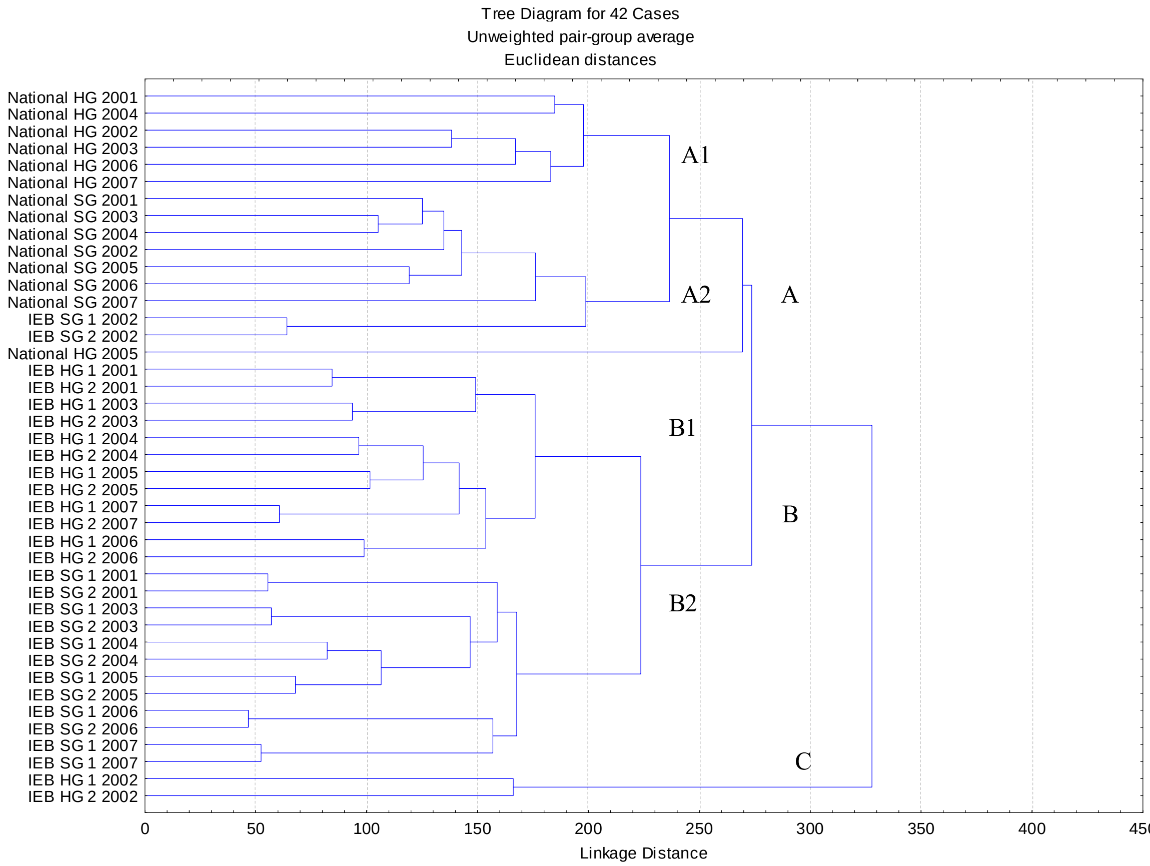


Figure 6.11 Results of a cluster analysis of the profiles of 2001-2007 national DoE and IEB SC Biology HG and SG question papers.

In the analyses of the CED-WCED-DoE HG question papers two groups of examinations, Groups A and B (Figure 6.9) were identified. Group A consisted of the 1994-2000 HG question papers which were set and administered by the CED and the WCED, using the CBS policy. Group B consisted of 2001-2007 HG question papers which were nationally set by the DoE, using the modified CBS policy, and administered by the WCED. Group A and Group B differed in that Group A question papers emphasized fewer different topics whereas Group B question papers emphasized more topics. Group A question papers separate as two Groups, A1 and A2 (Figure 6.9) based on the total marks of each question paper: A1 question papers were worth 400 marks each and A2 question papers each carried 300 marks. Both A1 and A2 question papers were each three hours long (Table 6.2) whereas the B question papers were four hours long (Table 6.3). The DoE 2005 HG examination (Figure 6.9 [B2]), was different to the other DoE HG examination (Figure 6.9[B1]), in that it included more short questions in Section A (Table 6.4), more questions of two-three sentences long, more questions which used non-text (Table 6.7), more short and extended answers (Table 6.14), and more Analyze questions (Table 6.19) than did the question papers in B2.

All IEB question papers were written over three hours. The IEB HG question papers showed two large groups, Group A and B (Figure 6.10). Group A1 question papers were similar to the Group B questions papers in that each was worth 400 marks and was written in three hours, but they had higher HOCS than the Group B examinations. A1 examinations did not offer a choice of questions in Section B of the question papers, and Section B carried more marks in 1994 and 1995 (Table 6.4). Group A2 consisted of question papers with a total of either 300 or 320 marks (Table 6.4) and these papers were relatively heterogeneous with respect to the structural aspects of the question papers (Tables 6.8, 6.9, 6.15 & 6.16). Group A2 question papers are grouped according to whether they have higher or lower HOCS (Appendices 6.51 - 6.54). Group B question papers were all worth 400 marks and differed from the 400-mark question papers in A1 in that they had lower HOCS (Appendices 6.49, 6.50 & 6.54). In group B1, the 2002 question paper profiles and the 1996 question paper profiles share similar use of non-text in questions (Tables 6.8 & 6.9), extended answers (Tables 6.15 & 6.16) and fewer HOCS than the profiles of the 1997 examination (Appendices 6.49, 6.50 & 6.54).

The 2001 to 2007 HG and SG question papers clustered into two larger Groups A and B generally represented the national DoE examinations and the IEB examinations, and Group C the two profiles of the IEB 2002 HG question papers (Figure 6.11). Group A consists of two smaller groups, namely, the national DoE HG examinations (Group A1) and the national DoE SG examinations (Group A2), which are distinguished from each other by both the number of marks per question paper and their respective DOK values. The IEB SG 2002 profiles clustered with the national DoE SG profiles because they carried total marks of 300 each. All the other IEB SG question papers were worth 225 marks each. Characteristics of the national DoE HG 2005 question paper that make it different from the other national DoE HG question papers have already been discussed (Table 6.19). The IEB question papers are divided into the profiles of the HG question papers (B1) each worth 330 marks, and the SG question papers (B2) each worth 225 marks. The 2002 HG question paper was unusual in that it carried 400 marks and hence was more like the IEB HG question papers of 1996 and 1997 (Figure 6.10).

The use of the absolute marks in comparisons of the profiles, rather than the use of marks transformed to some common mark (like a percentage), could be criticized on the basis that if a question paper had 400 marks all its attributes would be proportionally weighted more than similar attributes for a question paper worth 300 marks. Therefore, the profiles of 400-mark question papers might be expected to cluster together and the profiles of 300-mark question papers might be expected to cluster together. This pattern was not observed here – in both Figures 6.9 and 6.10, the profiles of question papers worth 400 marks formed two clusters (Figure 6.9, Group A1 and Group B; Figure 6.10, Group A1 and B) within each analysis. It might be argued that the national DoE policy with respect to topic weightings introduced post-2000 resulted in the DoE HG question papers grouping together (Figure 6.9, Group B). While this inference may be correct, the remaining question papers, with no pre-dictated topic weighting nonetheless cluster together in two distinct subgroups, namely Group A1 (1994 and 1995) (Figure 6.9), and Group A2 (1996 to 2000)(Figure 6.9). Groups A1 and A2 represent question papers with different totals but they also represent different examining bodies and therefore different examiners' interpretations of the CBS.

The exploratory cluster analyses described in this section suggest that the combination of attributes of question papers as measured and described in this study gives each question paper it own complex character or 'gestalt'[218] which possibly links to who the examiner(s) were and their particular interpretation of the prevailing examinations policy. Unfortunately there are no public records of who the examiners of each of the question papers analyzed in this study were, so this linkage could not be tested. The gestalt of questions papers is not obvious if the characteristics of question papers are described individually. Viewing the question papers in only one or two dimensions (content) only may also limits descriptions of this gestalt. Porter (2002) suggested that the two-dimensional language of content might benefit from additional dimensions, such as how the content is delivered to students, in future analyses. Further investigation, using different multivariate analytical approaches may tease out the relationships, if they exist, for example, between the attributes of question papers and content, which might increase the explanatory power of content, but this extension is beyond the scope of this study.

Analyses of the performance data from samples of students who wrote the 2005 and 2006 examinations provide some evidence of the way in which different attributes of various questions influence student performance. While the performance standards which are generated by the analysis of answer scripts depend on the content standards inherent in their corresponding question papers, the performance standards in turn ascribe meaning to the content standards because they explain the range of content standards which students have learned (Chapter 3).

[218] 'Gestalt' is used here as "[a]n integrated perceptual structure or unity conceived as functionally more than the sum of its parts" (Brown, 1993, p. 1082).

6.2 Analysis of answer scripts

Drawing on the distinction that Chomsky (1957) made between competence and performance, Messick (1994) noted that inferring students' competence from observations of performances or behaviour was not always easy to do, especially when links must be made between poor performance and a lack of competence. Constructs like relevant knowledge and skills, that is, content standards, support inferences made about student competency. In the SC Biology examinations, students were categorized and awarded a symbol which indicated a specific level of competence or mastery according to the aggregate mark which they achieved in the examination they wrote. In Chapter 3, an argument was made that the relevant knowledge and skills required to be learned and demonstrated by students be referred to as content standards, and that performance standards, defined as differential performance in specific content standards, indicate different levels of mastery or competence with respect to those content standards. The content standards, and therefore the performance standards, are influenced by structural aspects of examination question papers.

This section seeks to explicitly articulate the performance standards, for eight of the selected SC Biology examinations, (i.e., national DoE question papers, 2005 and 2006, HG and SG, Paper 1 and Paper 2) using the structural characteristics (Section 6.1.1) and the content standards (Section 6.1.2) generated for these SC question papers. Of necessity, some discussion of the difficulty of questions is touched on in this section but most of the discussion about question difficulty will only be addressed in Section 6.3.2.

The reader is reminded that for the SC Biology HG and SG examination marks, the cut-scores which separated different performance standards or performance levels, were preset and that they were the same for each year (Chapter 2, Figure 2.2). The different categories of performance standards were labelled by symbols A to H. Although the same symbols (i.e., A, [highest achievers, strongest students]) to H ([lowest achievers, weakest students]) were used by policy to label different performance levels in both the HG and SG examinations, each symbol was defined by different cut-scores in the HG and SG examinations, and the pass/fail cut-score was different for each of HG and SG (Chapter 2, Table 2.2). No performance level descriptions were given in the policy for what each of the recognized symbols mean in terms of different levels of mastery or competence. The section seeks to *ascertain* whether it is possible to create meaningful interpretations (i.e., performance standards) for each performance level and, if so, whether the performance level descriptions are the same between years, because the cut-scores were the same between years.

The cut-scores which separated the symbols from each other did not occur at fixed distances along the continuum of possible marks (Chapter 2, Table 2.2) which meant that graphs showing student performance along the continuum of symbols appears as shown in Figure 6.12. To simplify the presentation of student performance, and for ease of reading, the author chose to plot each graph with the symbols A to H as if they were equally spaced along the continuum of total marks. In each of the graphs, symbol A is preceded by the marks allocated to that variable on the question paper which is specified on each graph as question paper (QP) (Figure 6.12).

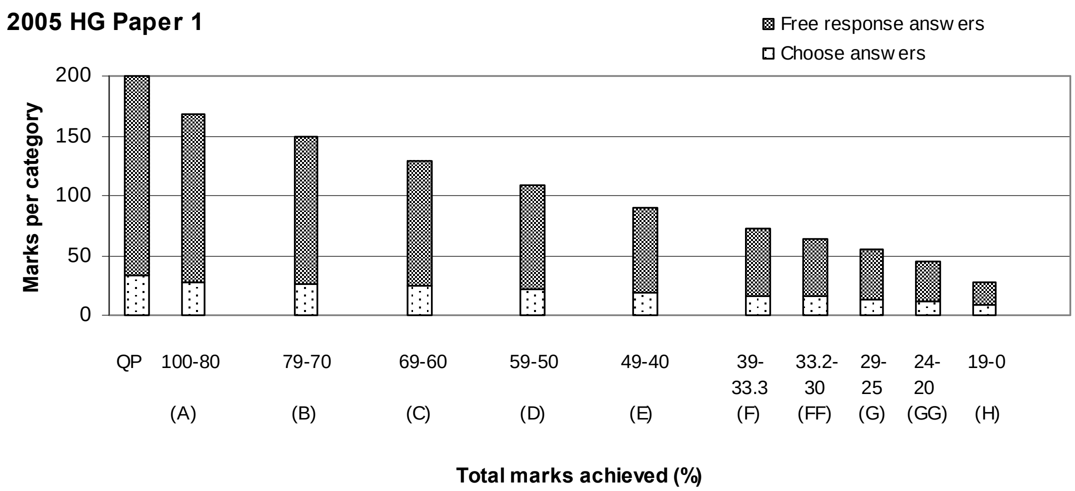


Figure 6.12 Student performance within answer type at different achievement levels. The achievement grade levels are shown as percentages with symbols in brackets. QP = question paper.

Questions which were deemed to be invalid at the subsequent memorandum discussion meetings (Chapter 5, Table 5.3) were excluded from the analyses of performance standards presented here, as all students were awarded the maximum marks for such questions and such questions therefore could not tell us anything about differences between categories of students. The DoE practice of awarding full marks for invalid questions results in an inappropriate inflation of student marks, and the reclassification of students into different categories of performance with invalid questions removed was presented in Chapter 5 (Tables 5.1 & 5.2).

6.2.1 Performance according to structure of question papers

Section 6.1.1 described and discussed particular structural features of SC Biology question papers thought to have had a possible influence on student performance. The current section examines whether each of these features affected student performance.

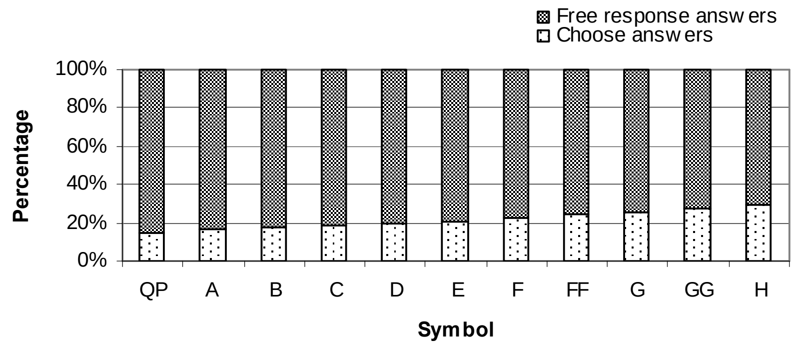
6.2.1.1 *Choosing the correct answers or free-response answers*

All eight SC Biology examinations show the same pattern: progressively students became weaker, they achieved more of their total marks from questions which required them to choose the correct answer from several possible correct answers (Figure 6.13). Possible explanations for this pattern are that weaker students guessed the answers from those given, or that it was easier for weaker students to recognize correct answers than to recall correct ones. This argument means that the proportion of marks allocated to either choosing the correct answer or requiring a free-response answer on a question paper would have influenced student performance on that question paper. While the proportion of each question paper which required free-response answers was similar in the eight HG and SG question papers for which there were student performance data, this measure was highly variable for the other years, for which there is no student performance data, and were different for the national DoE examinations and the IEB examinations (Section 6.1.1.2). This finding means that the proportion of free-response answers in question papers needs to be considered when setting question papers which are required to be equivalent.

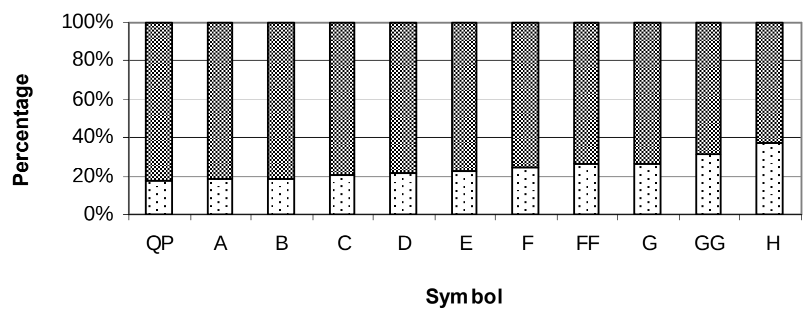
6.2.1.2 *Length of questions*

The length of questions influenced student performance differently and effects depended on the percentages for the various categories of question length on each question paper (Figure 6.14). For example, weaker students in the 2005 HG Paper 1 obtained more of their marks from shorter questions than from longer questions. This question paper was different to most other HG question papers (Section 6.1.3) because questions longer that three sentences almost always involved text together with a diagram, graph or a table or even different combinations of diagrams, graphs and tables. Weaker students might not have been able to process this combination of information in a question. Similarly, in the 2006 HG Paper 1, weaker students obtained more of their marks from shorter questions, rather than from longer questions which used text together with a diagram, graph or table. Conversely, weaker students in both 2005 SG question papers obtained more marks from longer questions. However, most of the 2005 SG longer questions included passages from which answers needed to be simply extracted verbatim from the text, or the question gave instructions about calculations to be made or graphs to be drawn.

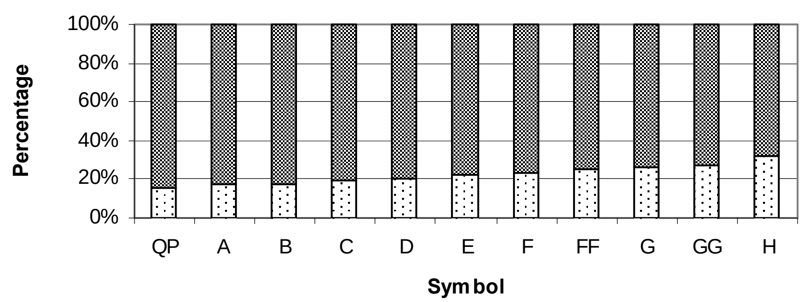
A. 2005 HG Paper 1



B. 2005 HG Paper 2



C. 2005 SG Paper 1



D. 2005 SG Paper 2

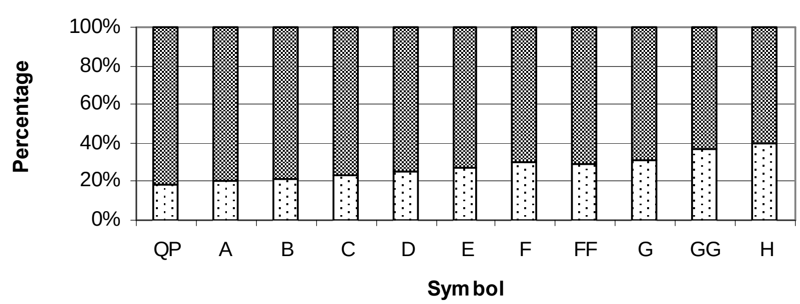
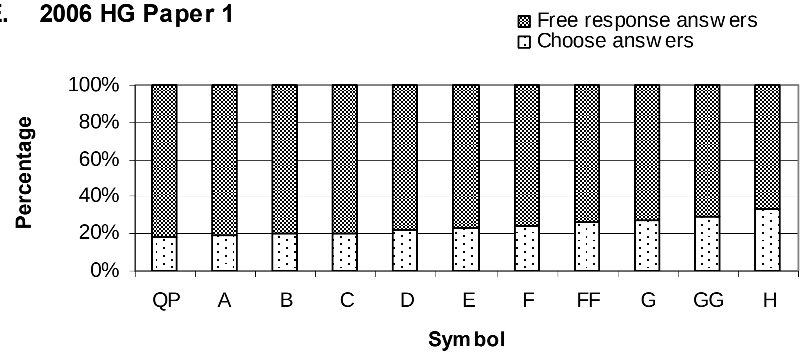
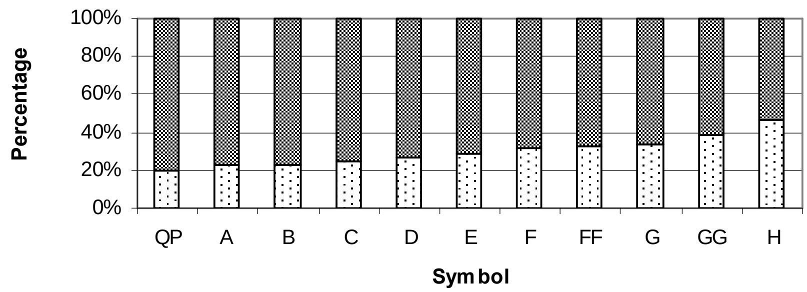


Figure 6.13 continued on next page

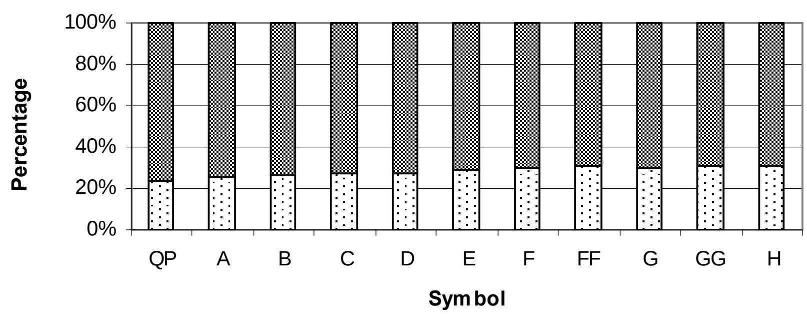
E. 2006 HG Paper 1



F. 2006 HG Paper 2



G. 2006 SG Paper 1



H. 2006 SG Paper 2

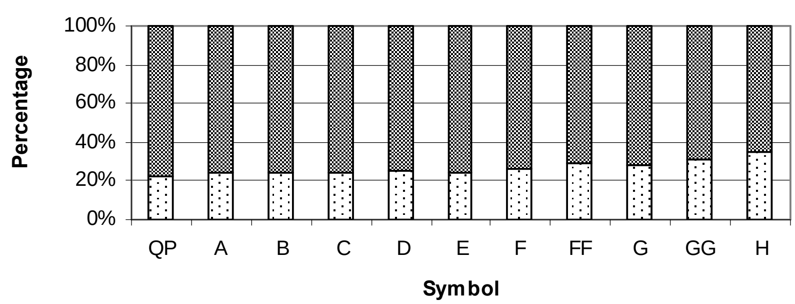
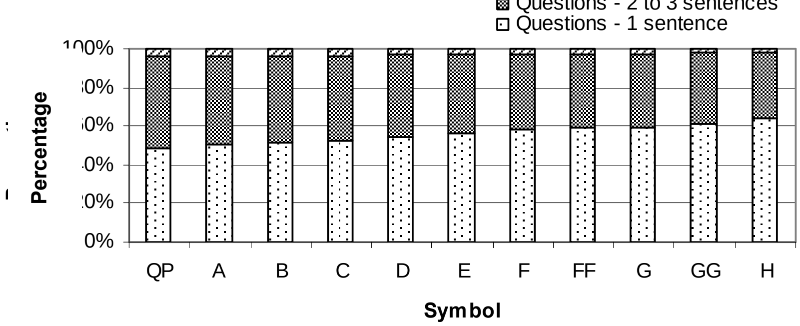
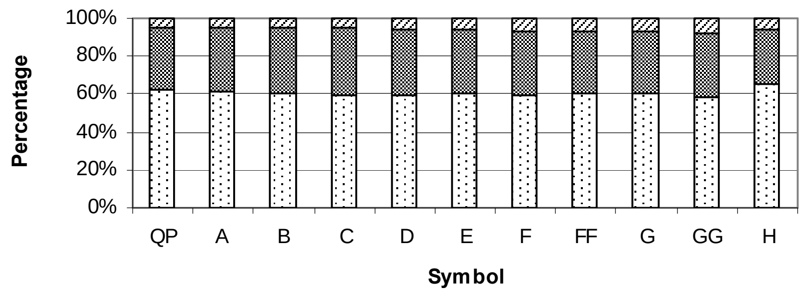


Figure 6.13 Student performance by answer type within achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2.

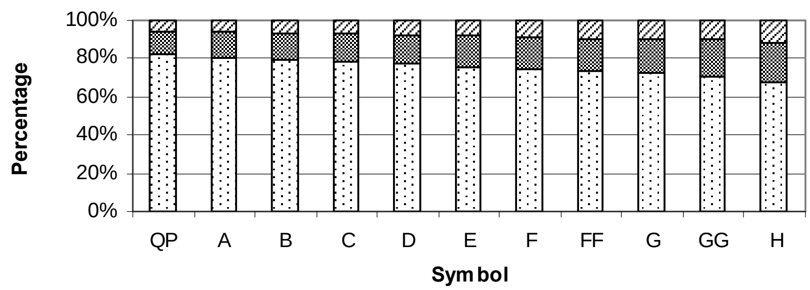
A. 2005 HG Paper 1



B. 2005 HG Paper 2



C. 2005 SG Paper 1



D. 2005 SG Paper 2

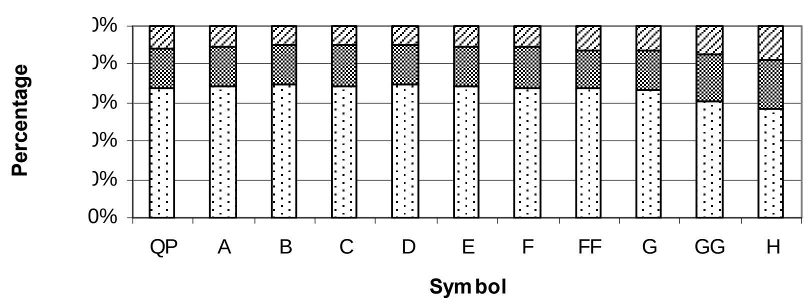
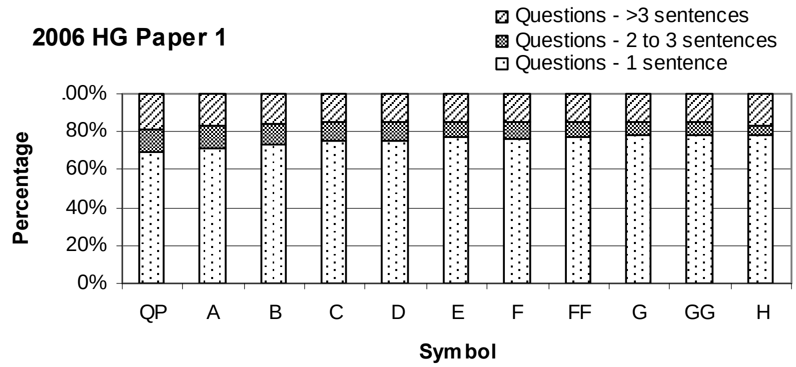
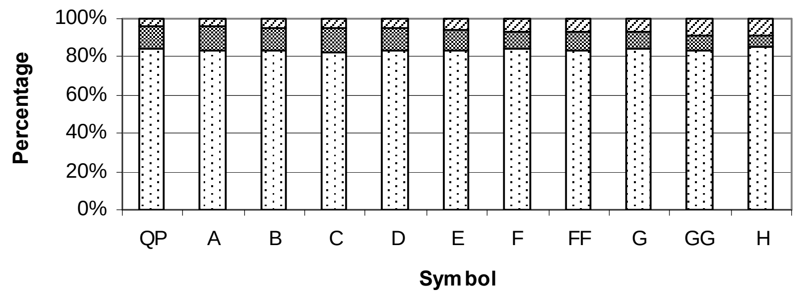


Figure 6.14 continued on next page

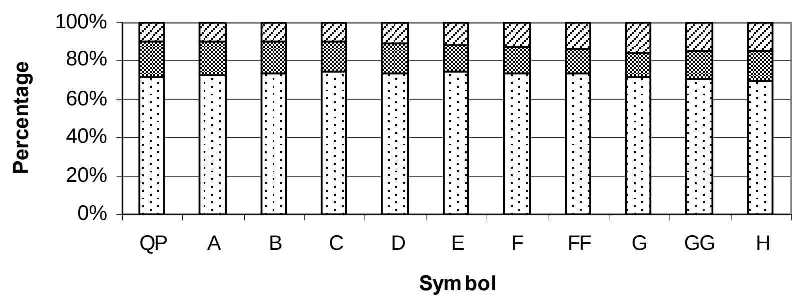
E. 2006 HG Paper 1



F. 2006 HG paper 2



G. 2006 SG Paper 1



H. 2006 SG Paper 2

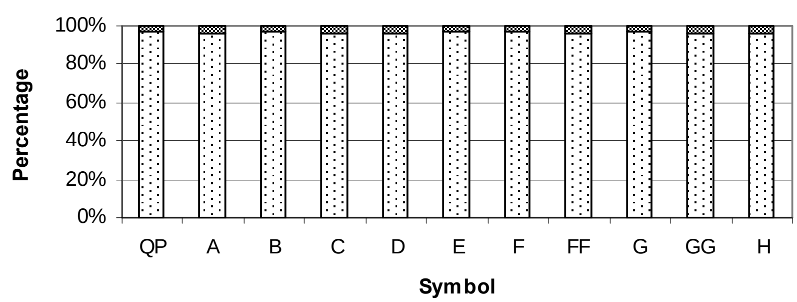


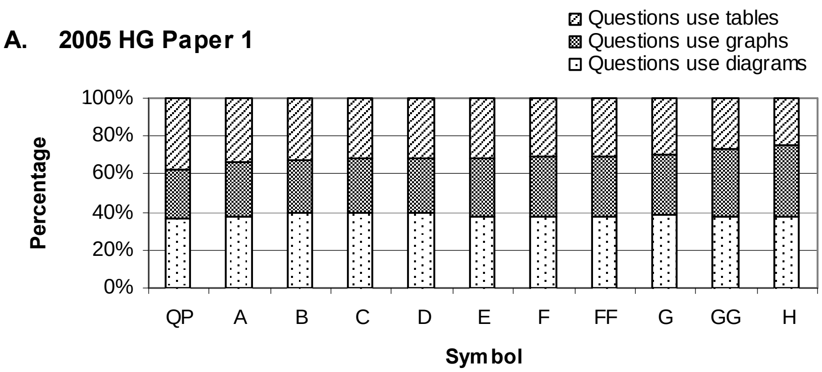
Figure 6.14 Student performance by question length within achievement levels , DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2.

As noted previously (Section 6.1.1.2a), the length of questions as defined and measured in this research may not be the optimal way to investigate the impact of reading a question on how students answer that question. For example, post-2000 many of the SG question papers with questions of more than one sentence long had very short sentences compared to the length of the sentences used in the HG question papers. Given the important role of reading and writing in academic literacy (Yeld, 2009), these aspects of how SC examination questions are structured and how structure impacts upon student performance deserve more work in the future, especially in South Africa where many SC students write their examinations in a language which is not their home language.

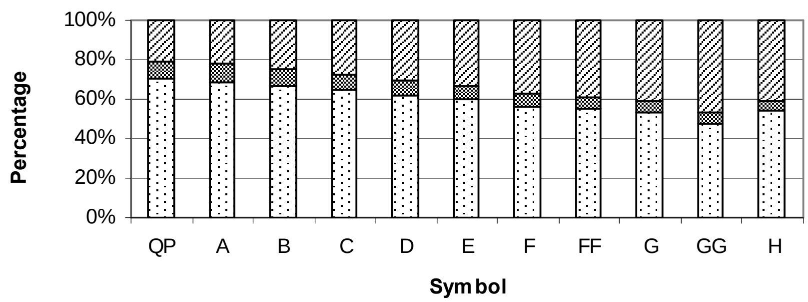
6.2.1.3 *Use of non-text elements in questions*

The use of non-text elements, that is, tables, graphs and diagrams, in examination questions influenced student performance differently in different question papers, and depended on the proportion of each of the different categories, or combinations of categories, of non-text in each question paper (Figure 6.15). For example, when the students needed to apply information using the non-text elements, or put together information presented in two different non-text formats, (e.g., 2005 HG Paper 1 and Paper 2), higher achieving students could do these tasks better than lower achieving students. Where non-text information presented in the question mimicked what was specified in the CBS, there was little difference in how the relative performance of the range of students in the sample. The 2006 Paper 1 and Paper 2 scripts came from a common sample of students for each of HG and SG. Weaker HG students were better able to better answer questions which involved the interpretation of tables than those questions which involved the interpretation of diagrams. Weaker students in the 2006 HG Paper 1 performed relatively poorly in comparison to the more able students in a question which involved the interpretation of a chromatograph─an applied question because it was not required by the syllabus. Students had to understand what chromatography was from text and a diagram.

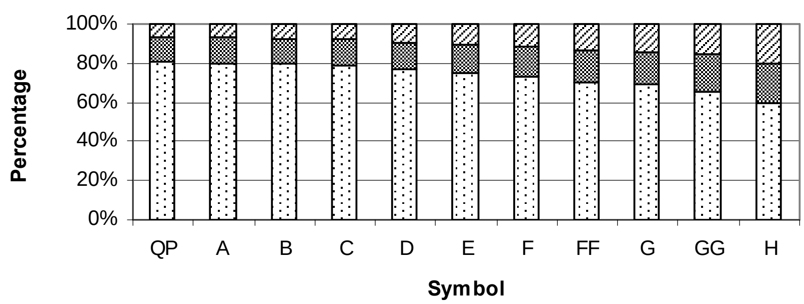
A. **2005 HG Paper 1**



B. **2005 HG Paper 2**



C. **2005 SG Paper 1**



D. **2005 SG Paper 2**

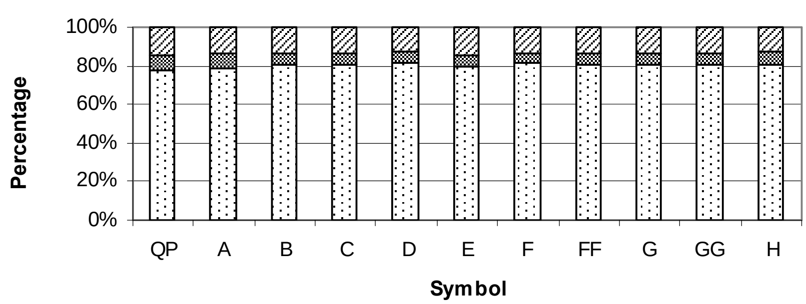
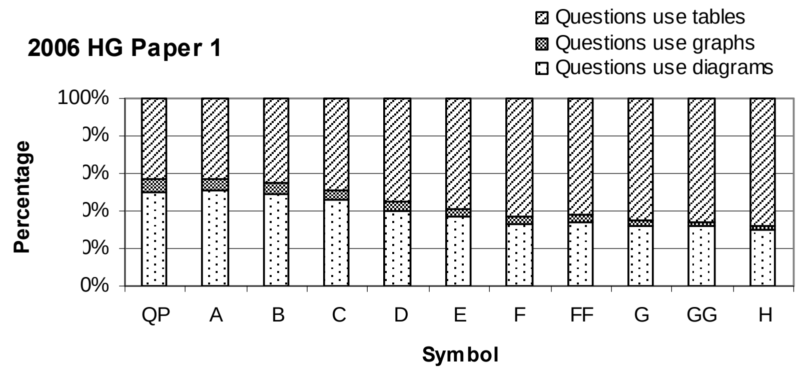
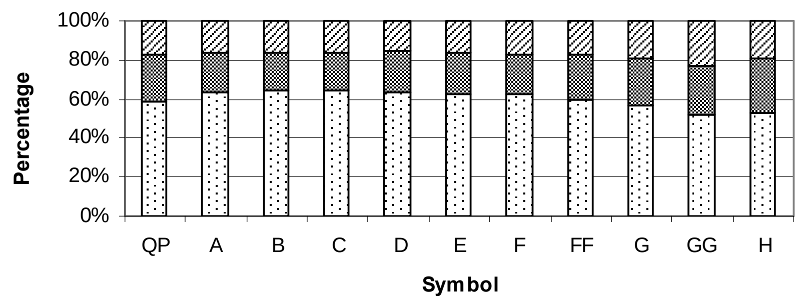


Figure 6.15 continued on next page

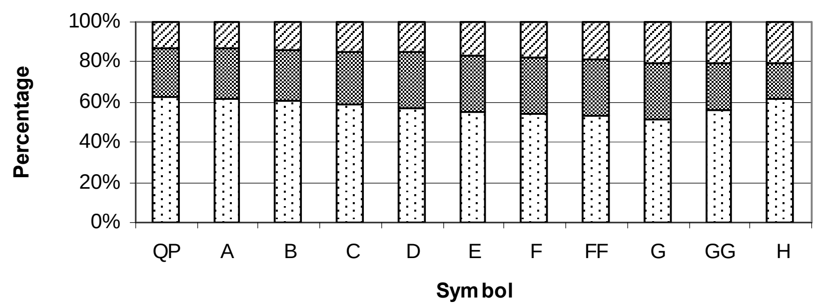
E. 2006 HG Paper 1



F. 2006 HG paper 2



G. 2006 SG Paper 1



H. 2006 SG Paper 2

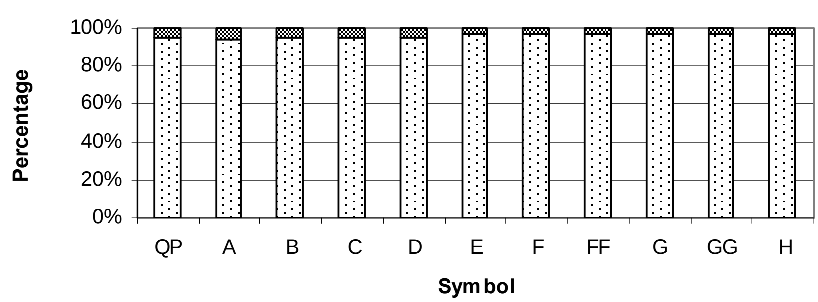


Figure 6.15 Student performance by non-text elements in questions within achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2.

6.2.1.4 *Length of answers*

The performance of the spectrum of students for all eight questions papers with respect to length of answers was similar − weaker students consistently answered correctly relatively fewer extended answer questions than questions with a short answer of one to two term answers, irrespective of the relative proportions for each of the different categories of answer length on each question paper (Figure 6.16). This outcome was surprising, given that many longer answers did not require any logical order (Section 6.1.1.2a) or evidence of understanding. This suggests that some longer recall answers might be more difficult than shorter recall answers. This contrast will be further discussed in the section on difficulty (Section 6.3.2).

6.2.1.5 *Non-text elements required in answers*

Generally, in 2005 HG Paper 1 and Paper 2 stronger students were more often able to produce tables, diagrams and graphs when required to do so by questions, in contrast to the weaker students (Figure 6.17). Weaker 2005 HG students had difficulty producing diagrams as answers (Figure 6.17), but could construct tables when required to do so (Figure 6.17). By contrast weaker students in the 2006 HG Paper 1 were relatively weaker in composing tables. The difference between the performance of the weaker students in the 2005 and 2006 HG Paper 1 versions was possibly due to the fact that to answer with the table required in 2005 students needed to tabulate the results of an iodine test on three leaves, no matching of facts were required. The 2006 table required a tabulation of three differences to be made between photosynthesis and cellular respiration – here students had to match three points of difference. In some question papers, weaker students were better able to draw the required table than the required diagram (e.g., 2005 HG Paper 1) or were less able to draw the required table than the required graph or diagram (e.g., 2006 HG Paper 1). Clearly whether students can draw tables, graphs or diagrams depends on the context of each non-visual answer, and whether students needed to translate between verbal and non-verbal aspects of a question.

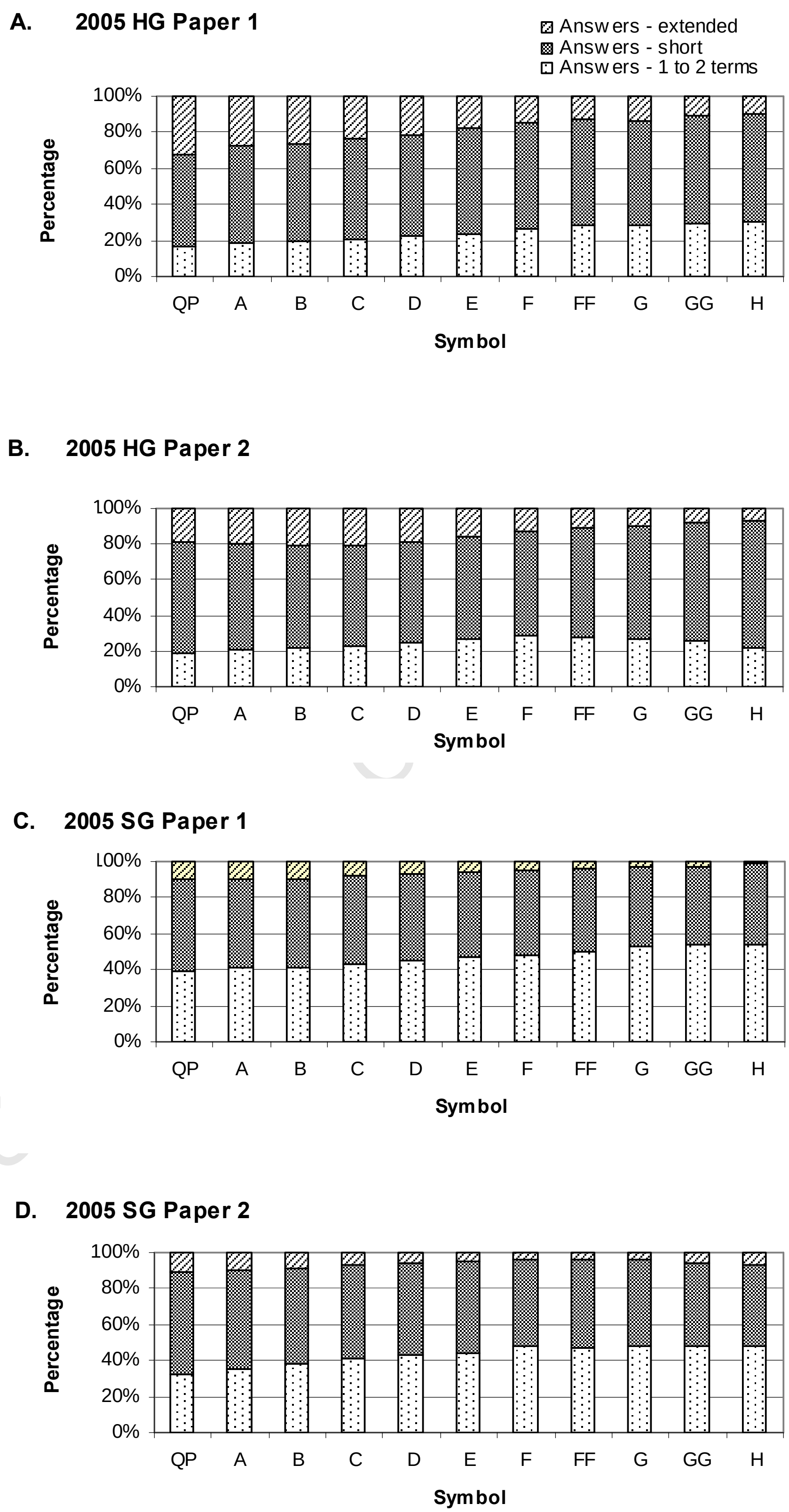
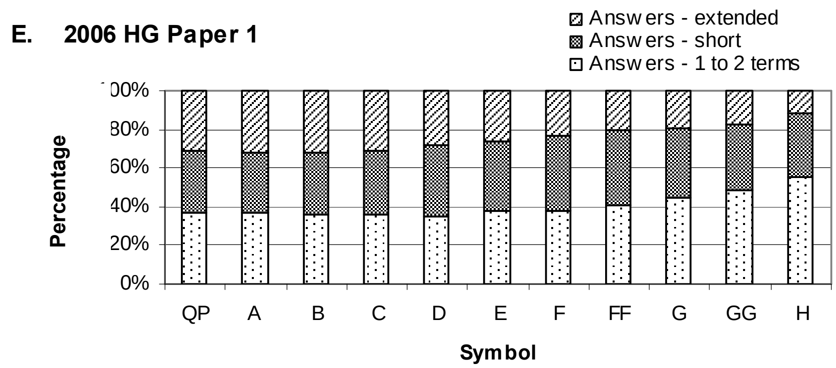
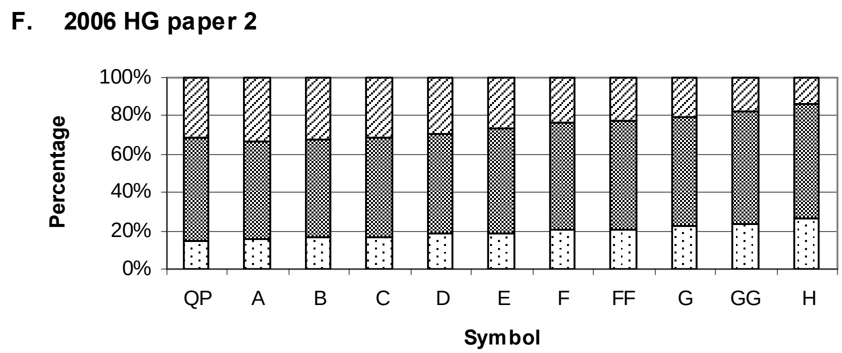
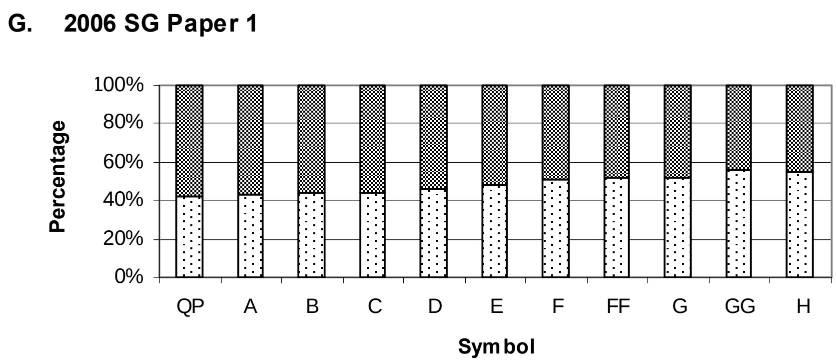


Figure 6.16 continued on next page







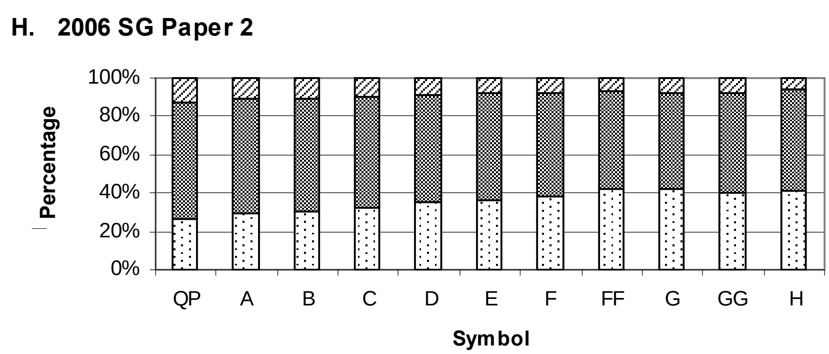
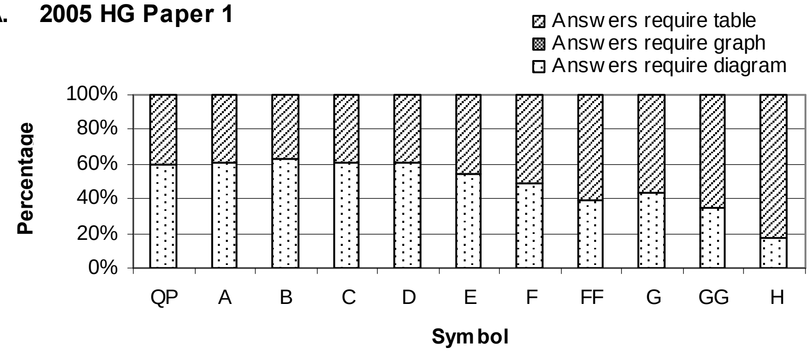
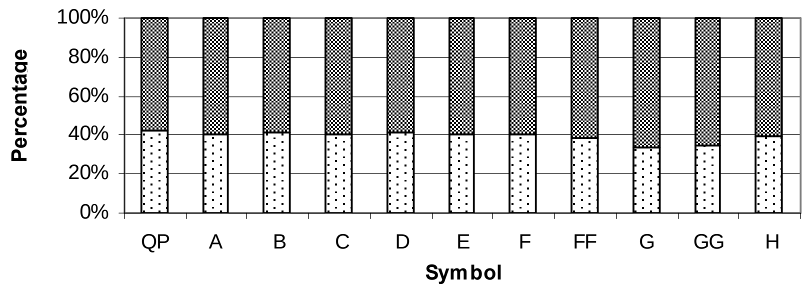


Figure 6.16 Student performance by the length of answers required within achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2.

A. 2005 HG Paper 1



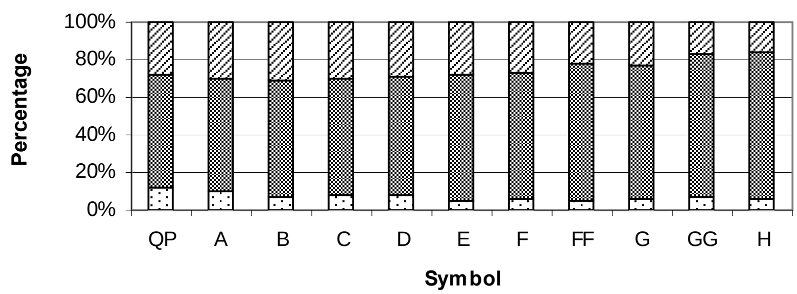
B. 2005 HG Paper 2



C. 2005 SG Paper 1 – answers require no tables, graphs or diagrams

D. 2005 SG Paper 2 – answers require diagram only

E. 2006 HG Paper 1



F. 2006 HG Paper 2 – answers require graph only

G. 2006 SG Paper 1 – answers require diagram only

Figure 6.17 continued on next page

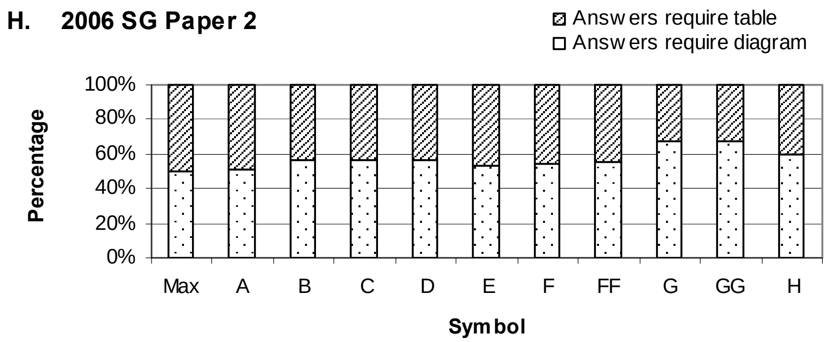


Figure 6.17 Student performance by answers which require non-text elements within achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2.

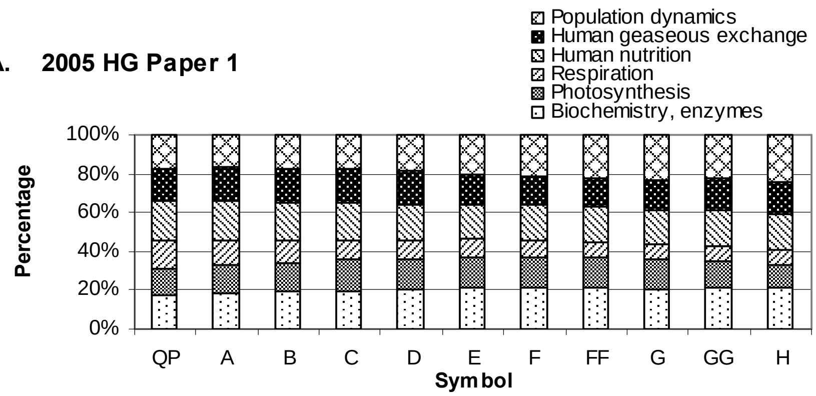
6.2.2 Performance standards

Performance standards recognize categories of proficiency, competency or mastery achieved in particular content standards. There are five possible ways or constructs by which one could view performance standards in terms of the content standards examined in an examination. These ways are as individual marks obtained in each of the scorable events of the examination; as summed marks obtained for each of the topics; as summed marks obtained for each of the performance expectations; as summed marks obtained for each combination of topic and performance expectation, or as a total mark obtained in the examination. Each of these different ways of viewing performance standards will be explored in the following discussion.

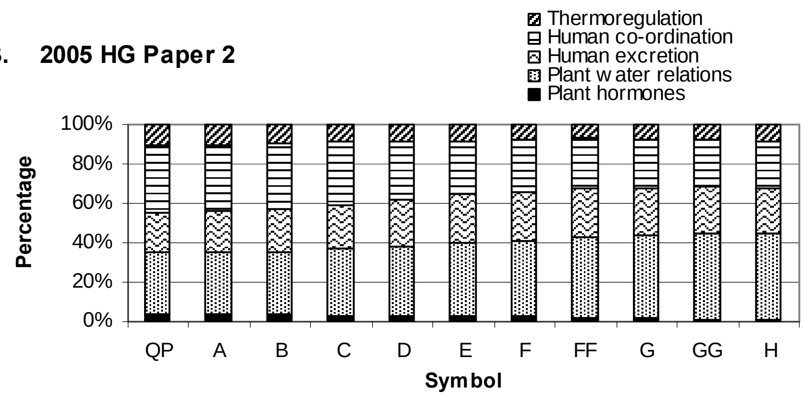
6.2.2.1 *Performance according to topics*

Generally, student performance according to topics varied, depending on the relative emphases of topics in each question paper. Exceptions were, for example, in three of the four question papers which included questions about population dynamics (2005 HG Paper 1, 2005 SG Paper 1, 2006 SG Paper 1), weaker students performed better in this topic than they did in other topics (Figure 6.18 [[A,C,G]). The population dynamics questions in the 2006 HG Paper 1 required students to read a passage of text about an ecosystem and answer conceptual questions about this ‘new’ ecosystem, weaker students were less successful in answering these questions (Figure 6.18 [E]). There were no obvious topics in which some categories of students consistently performed relatively better than in other topics.

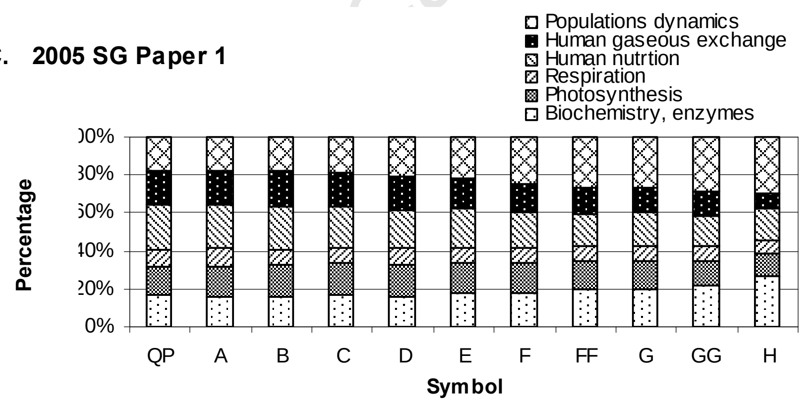
A. 2005 HG Paper 1



B. 2005 HG Paper 2



C. 2005 SG Paper 1



D. 2005 SG Paper 2

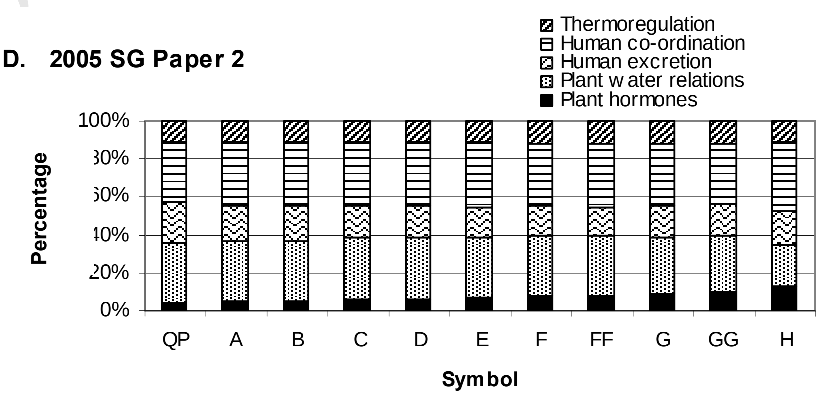
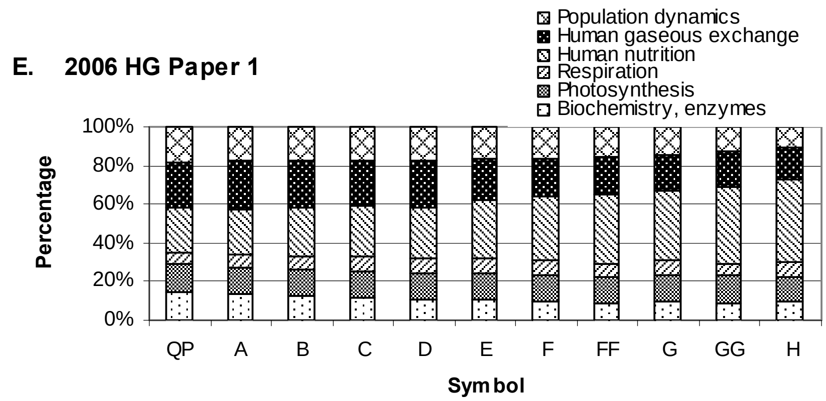
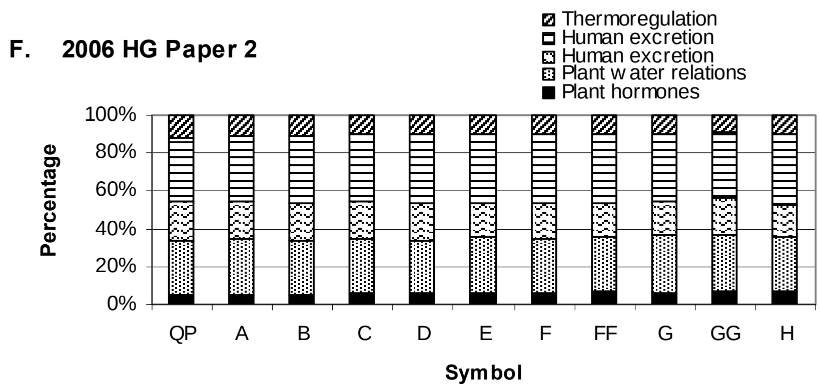
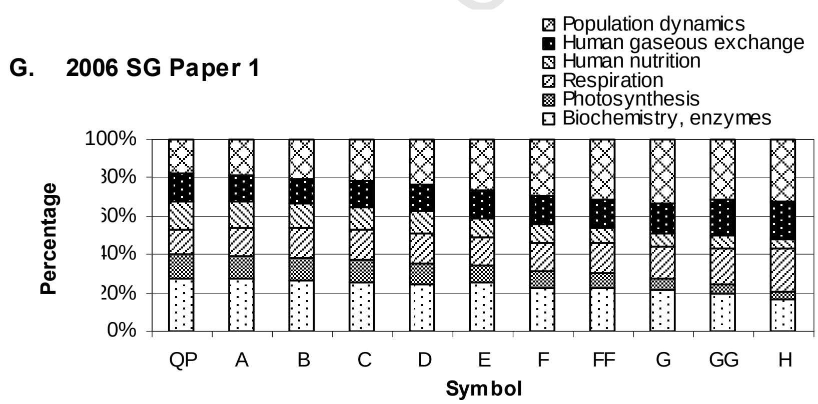


Figure 6.l8 continued on next page







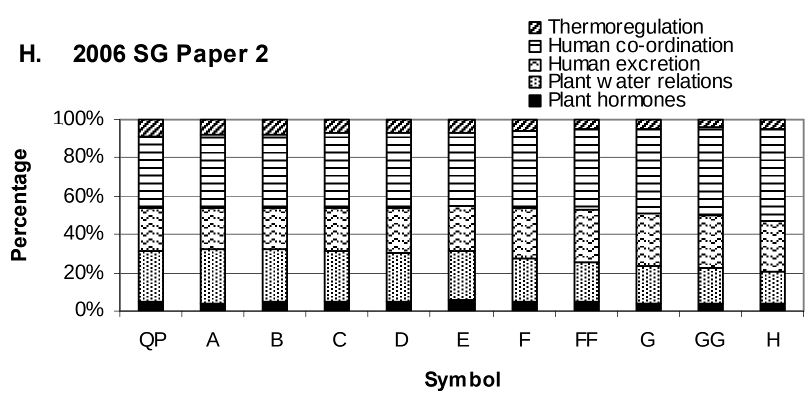


Figure 6.18 Student performance by topics within achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2.

For each of the eight question papers, student performance in each of the categories of topics were all significantly correlated with each other and with the total marks that students obtained in each question paper (Tables 6.29 & 6.30). In Section 6.1.2.1 it was hypothesized that the presence of questions about biochemistry, photosynthesis, respiration and water relations, because these topics involved abstract understanding, might have had a negative effected student performance. There is some evidence, for example, in 2006 SG Paper 1 which contained four of the five possibly more difficult topics, that this might have occurred. This issue will be discussed again in the Section 6.3.

6.2.2.2 *Performance according to performance expectations*

Higher order cognitive demand is increasingly being implicated as important in how individuals are taught and what they learn (e.g., Alberts, 2009, 2010), but not all assessments offer students the opportunity to demonstrate their competence in HOCS (e.g., Zheng et al., 2008). In South Africa, there have been repeated calls for examinations to test HOCS (e.g., Reddy, 2006a), but there is no common understanding of what HOCS means across different subjects (e.g., Crowe, 2010; Muller, 2005; Umalusi, 2007). This has led to the development of the PET used in this study (Chapter 4).

The extent of HOCS recorded in the SC Biology examinations investigated in this study varied considerably between different examinations (Section 6.1.2.2) and determining how much HOCS a SC Biology question paper should test is beyond the scope of this thesis. What this section thus seeks to do, is to examine what differences, if any, existed in the performance of students with respect to HOCS (and LOCS) in eight different SC examinations, and the possible implications thereof for teaching and learning Biology, given international calls for greater emphases on HOCS in teaching, learning and assessment.

Student performance in LOCS and HOCS across the achievement symbols for each question paper was remarkably similar for each of the different PET categories of each question paper (Figure 6.19). That is, all students irrespective of whether they are top achievers or low achievers demonstrated both LOCS (i.e., Memorize, Perform-Routine-Procedures, Explain) and HOCS (i.e., Analyze, Apply) in similar proportions to each other and to that of the question

Table 6.29 Correlations between the performances of candidates in the 2005 WCED examinations with respect to topic within each of the Papers 1 and 2. Different candidates constitute the sample for each of Paper 1 and Paper 2 and for each of HG and SG.

HG (n = 1001)

Paper 1	Biochemistry	Photosynthesis	Respiration	Human nutrition	Human gaseous exchange	Population dynamics	Total mark
Biochemistry	1.00	*0.85*	*0.77*	*0.87*	*0.81*	*0.79*	***0.93***
Photosynthesis		1.00	*0.76*	*0.84*	*0.81*	*0.77*	***0.92***
Respiration			1.00	*0.79*	*0.79*	*0.74*	***0.88***
Human nutrition				1.00	*0.84*	*0.79*	***0.94***
Human gaseous exchange					1.00	*0.76*	***0.92***
Population dynamics						1.00	***0.88***

SG (n = 982)

Paper 1	Biochemistry	Photosynthesis	Respiration	Human nutrition	Human gaseous exchange	Population dynamics	Total mark
Biochemistry	1.00	*0.70*	*0.65*	*0.74*	*0.42*	*0.68*	***0.84***
Photosynthesis		1.00	*0.75*	*0.83*	*0.83*	*0.69*	***0.91***
Respiration			1.00	*0.77*	*0.76*	*0.64*	***0.84***
Human nutrition				1.00	*0.80*	*0.76*	***0.94***
Human gaseous exchange					1.00	*0.70*	***0.93***
Population dynamics						1.00	***0.82***

HG (n = 1004)

Paper 2	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermoregulation	Total mark
Plant hormones	1.00	*0.77*	*0.72*	*0.77*	*0.74*	***0.82***
Plant water relations		1.00	*0.84*	*0.87*	*0.82*	***0.95***
Human excretion			1.00	*0.84*	*0.78*	***0.92***
Human co-ordination				1.00	*0.85*	***0.96***
Thermoregulation					1.00	***0.89***

SG (n = 854)

Paper 2	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermoregulation	Total mark
Plant hormones	1.00	*0.47*	*0.42*	*0.48*	*0.41*	***0.55***
Plant water relations		1.00	*0.77*	*0.79*	*0.41*	***0.93***
Human excretion			1.00	*0.80*	*0.68*	***0.89***
Human co-ordination				1.00	*0.72*	***0.94***
Thermoregulation					1.00	***0.82***

Table 6.30 Correlations between the performance of candidates in the 2006 WCED examinations with respect to topics within each of the Papers 1 and 2. The same candidates constitute the samples for both Paper 1 and Paper 2 within each of HG and SG.

	Paper 1	Biochemistry	Photosynthesis	Respiration	Human nutrition	Human gaseous exchange	Population dynamics	Total mark
HG (n = 975)	**Biochemistry**	1.00	*0.80*	*0.72*	*0.77*	*0.80*	*0.77*	***0.90***
	Photosynthesis		1.00	*0.77*	*0.78*	*0.81*	*0.77*	***0.90***
	Respiration			1.00	*0.72*	*0.74*	*0.70*	***0.82***
	Human nutrition				1.00	*0.81*	*0.79*	***0.91***
	Human gaseous exchange					1.00	*0.80*	***0.94***
	Population dynamics						1.00	***0.90***
SG (n = 881)	**Biochemistry**	1.00	*0.76*	*0.77*	*0.80*	*0.72*	*0.69*	***0.93***
	Photosynthesis		1.00	*0.73*	*0.74*	*0.69*	*0.63*	***0.87***
	Respiration			1.00	*0.74*	*0.71*	*0.65*	***0.87***
	Human nutrition				1.00	*0.75*	*0.65*	***0.89***
	Human gaseous exchange					1.00	*0.61*	***0.84***
	Population dynamics						1.00	***0.80***

	Paper 2	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermoregulation	Total mark
HG (n = 975)	**Plant hormones**	1.00	*0.67*	*0.64*	*0.69*	*0.64*	***0.74***
	Plant water relations		1.00	*0.86*	*0.88*	*0.81*	***0.95***
	Human excretion			1.00	*0.87*	*0.81*	***0.93***
	Human co-ordination				1.00	*0.85*	***0.97***
	Thermoregulation					1.00	***0.89***
SG (n = 881)	**Plant hormones**	1.00	*0.65*	*0.58*	*0.62*	*0.65*	***0.70***
	Plant water relations		1.00	*0.79*	*0.83*	*0.73*	***0.94***
	Human excretion			1.00	*0.79*	*0.69*	***0.90***
	Human co-ordination				1.00	*0.71*	***0.95***
	Thermoregulation					1.00	***0.80***

paper.[219] Top-achievers showed evidence that they could successfully answer both more LOCS and HOCS questions that the lower-achieving students could. In this study, an explicit conceptual link is made between the categories of cognitive levels that characterize LOCS and the categories of cognitive demand that characterize HOCS (Chapter 4).

Results presented in Figure 6.19 suggest that if students understood what they memorized and the procedures that they routinely performed, then there was a greater likelihood that they used this knowledge productively in questions which required HOCS, a view also espoused by Sousa (2006).[220] The inability of students from the USA to compete in international tests testing HOCS was attributed to "expecting critical and advanced thinking skills from kids who have been trained to regard facts and substantive knowledge as unimportant" (Booker, 2007, p.347). Other researchers, for example, Chang (2010), showed that the importance of prior or learned knowledge (LOCS) was different for different categories of problem-solving or HOCS.

This study suggests that the focus in teaching Biology at SC level should be on ensuring that students have an understanding of the facts and concepts which they are required to learn, and in consequence could successfully complete more HOCS questions. A further implication is that tasks such as examination questions which require students to demonstrate understanding should also require that some logical organization of the answer be provided, thus discouraging the use of 'any x facts [in any order]' in memoranda as discussed above (Section 6.1.1.2a)

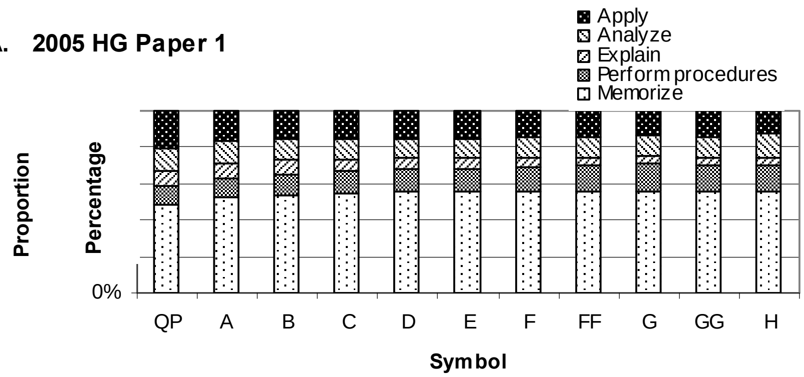
Given how LOCS and HOCS were defined in this study, in many learning situations tasks are considered HOCS because they are new to students, and once learned the same tasks are recognized as LOCS for assessment purposes. In acquiring the Explain category of LOCS (i.e., after they are *first* taught the facts, concepts, processes or procedures), students are practicing HOCS as defined by this study, provided they are not simply repeating an explanation given to them by their teacher.[221] Sousa (2006) and Marshall and Horton (2011) promoted a similar idea, namely, that achieving understanding through the exploration of the meaning of concepts and procedures better equips students to move onto deeper levels of understanding.

219 Data from this study would support Sousa's (2006) argument that the difficulty of a task rather than its complexity (LOCS less complex, HOCS more complex) relates to student ability. Sousa (2006, p. 258) described complexity as "the *thought process* that the brain uses to deal with information" and difficulty as the "*amount of effort* that the learner must expend *within* a level of complexity to accomplish a learning objective. It is possible for a learning activity to become increasingly difficult without becoming more complex".

220 Dewey (1938, p.79) offered a similar view in the language of his time. "First that the problem grows out of the conditions of the experience being had in the present, and that it is within the range of the capacity of the students; and, secondly, that it is such that it arouses in the learner an active quest for information and for the production of new ideas. The new facts and ideas thus obtained become the ground for further experiences in which new problems are presented".

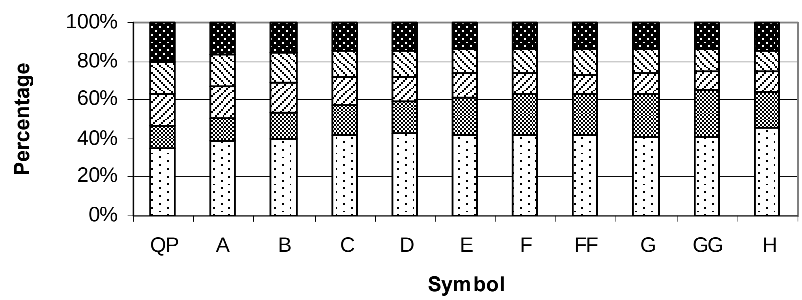
221 Unfortunately this study had no way of discerning whether answers given by students at this level in the SC examinations were really understood by the student or not.

A. 2005 HG Paper 1

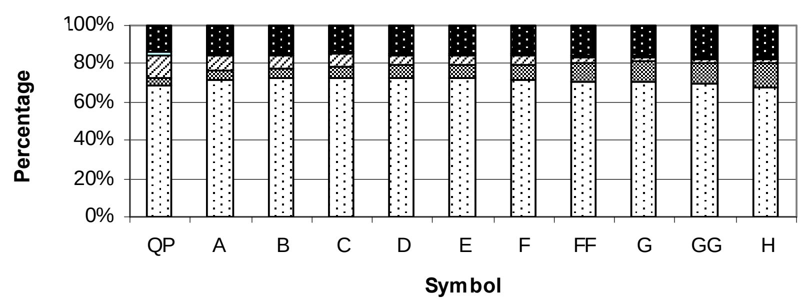


Sousa

B. 2005 HG Paper 2



C. 2005 SG Paper 1



D. 2005 SG Paper 2

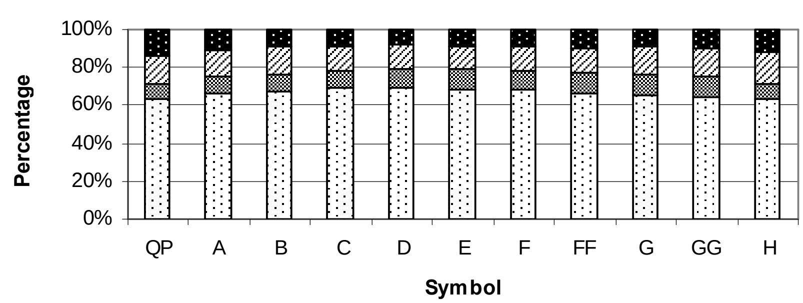
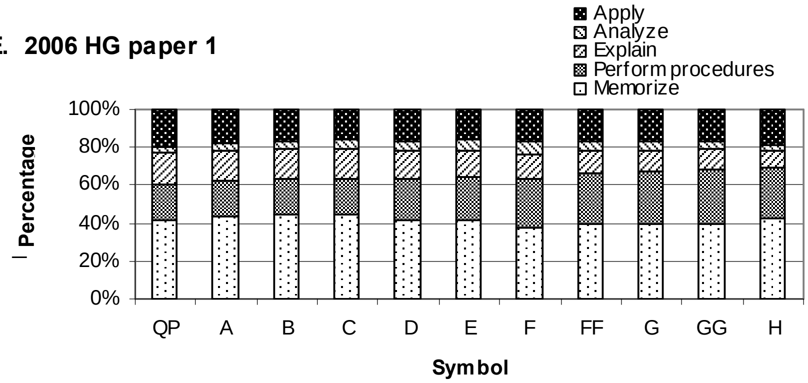
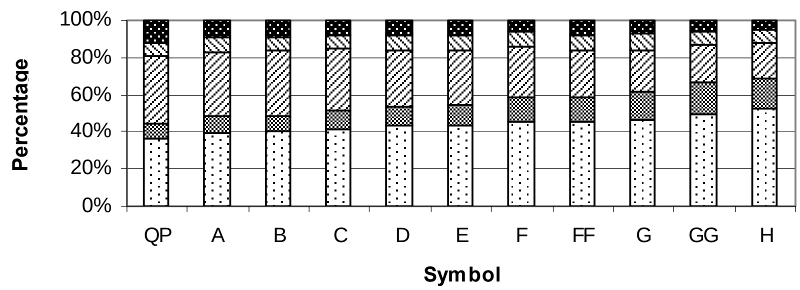


Figure 6.19 continued on next page

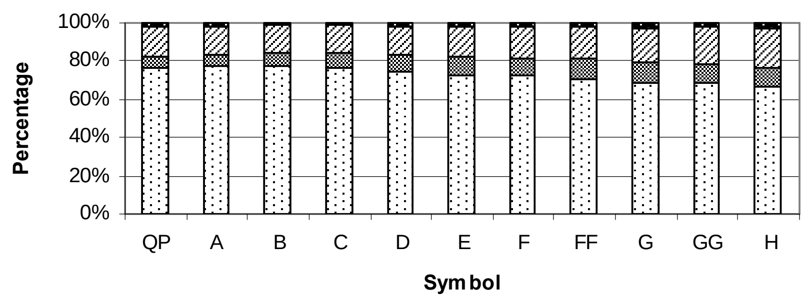
E. 2006 HG paper 1



F. 2006 HG Paper 2



G. 2006 SG Paper 1



H. 2006 SG Paper 2

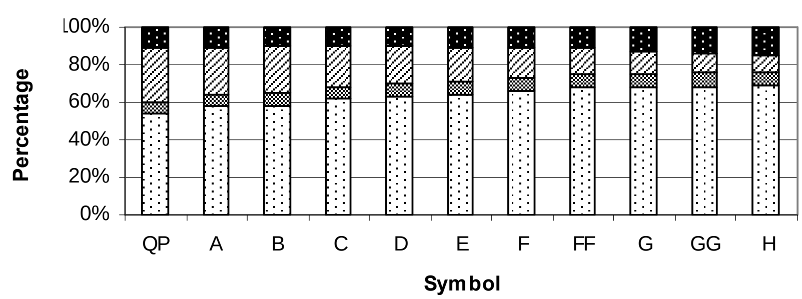


Figure 6.19 Student performance by PET categories with achievement levels, DoE 2005 and 2006, HG Paper 1 and Paper 2, and SG Paper 1 and Paper 2.

Others such as Haak et al. (2011, p. 1215) argued differently in that “active learning does not help with information transfer, only with problem-solving and other types of higher order learning”. To support this assertion Haak et al. (2011, p.1215) stated that “active-learning exercises have been shown to increase performance on exam questions that demand higher-order cognitive skills, while having no effect on exam questions focused on lower-order cognitive skills [levels 3 to 6 versus 1 and 2 on Bloom’s taxonomy; e.g., (Lord, 1997)]. These authors misrepresented Lord (1997, p. 214) who stated that “the study found that by striving for student understanding rather than rote memorization, the experimental students [in a student-centered learning environment] performed significantly better th[a]n their traditionally taught counterparts [in a teacher-centered learning environment]”. Haak et. al (2011) viewed comprehension as a part of LOCS and Lord (1997) recognized two categories: memorization/rote and the understanding of material. This misrepresentation highlights the danger of moving too rapidly between different conceptualizations of categories of cognitive demand in research (see Chapter 4). Lord’s (1997) view supports the transitional importance ascribed in this study to the LOCS category ‘explain’ in that it connects memorized facts, concepts, and routine procedures to HOCS, which represent deeper levels of understanding (Chapter 4). Do HOCS not involve the transfer and re-arrangement of the knowledge of facts, concepts and procedures (i.e., LOCS) in ways that are new and novel to the person engaged in HOCS activity? According to Sousa (2006) they do, and the PET explicitly links LOCS and HOCS. Sousa (2006) argued that teachers needed to devote sufficient time for students to process their repertoire of knowledge so that they could use it to address tasks that required HOCS thinking.

“[T]o comprehend science as a responsible citizen, and certainly to succeed in any science-related career, *both* content and reasoning are essential” (Gross, 2009, p. 35) and “domain knowledge strongly effects the quality of thinking” (Gross, 2009, p. 39). The research reported in this thesis suggests that performance in LOCS and HOCS are clearly related. For each of the eight question papers analyzed here student performance in the categories of PET were all significantly correlated with each other and with the total marks that students obtained (Tables 6.31 & 6.32). In addition, the 2006 performance of the same students in each of PET categories and the total mark in Paper 1 are correlated to their performance in the same PET categories and the total mark achieved in Paper 2 (Table 6.33). This pervasive correlation suggests that particular categories of PET in isolation do not aid discrimination between students of different abilities.

Table 6.31 Correlations between the cognitive demand performance of candidates in the 2005 WCED examinations within each of the Papers 1 and 2. Different candidates form the sample for each of Paper 1and Paper 2 for each of HG and SG.

	Paper 1	Memorize	Procedure	Explain	Analyze	Apply	Total mark
HG (n = 1001)	**Memorize**	1.00	*0.82*	*0.87*	*0.83*	*0.86*	***0.98***
	Procedure		1.00	*0.75*	*0.74*	*0.77*	***0.87***
	Explain			1.00	*0.81*	*0.81*	***0.90***
	Analyze				1.00	*0.82*	***0.89***
	Apply					1.00	***0.92***
SG (n = 982)	**Memorize**	1.00	*0.48*	*0.85*	*0.53*	*0.73*	***0.99***
	Procedure		1.00	*0.41*	*0.26*	*0.47*	***0.53***
	Explain			1.00	*0.47*	*0.67*	***0.88***
	Analyze				1.00	*0.45*	***0.55***
	Apply					1.00	***0.81***

	Paper 2	Memorize	Procedure	Explain	Analyze	Apply	Total mark
HG (n = 1004)	**Memorize**	1.00	*0.74*	*0.90*	*0.87*	*0.84*	***0.97***
	Procedure		1.00	*0.69*	*0.69*	*0.65*	***0.79***
	Explain			1.00	*0.86*	*0.84*	***0.94***
	Analyze				1.00	*0.86*	***0.93***
	Apply					1.00	***0.91***
SG (n = 854)	**Memorize**	1.00	*0.63*	*0.77*	•	*0.65*	***0.98***
	Procedure		1.00	*0.53*	•	*0.47*	***0.71***
	Explain			1.00	•	*0.60*	***0.84***
	Analyze				•	•	•
	Apply					1.00	***0.73***

Table 6.32 Correlations between the cognitive demand performance of candidates in the 2006 WCED examinations with respect to cognitive demand within each of the Papers 1 and 2. The same candidates form the samples for both Paper 1 and Paper 2 for each of HG and SG.

	Paper 1	**Memorize**	**Procedure**	**Explain**	**Analyze**	**Apply**	**Total mark**
HG (n = 975)	**Memorize**	1.00	*0.73*	*0.86*	*0.64*	*0.81*	***0.96***
	Procedure		1.00	*0.75*	*0.54*	*0.76*	***0.85***
	Explain			1.00	*0.60*	*0.81*	***0.93***
	Analyze				1.00	*0.55*	***0.68***
	Apply					1.00	***0.90***
SG (n = 881)	**Memorize**	1.00	*0.63*	*0.76*	•	*0.43*	***0.99***
	Procedure		1.00	*0.60*	•	*0.32*	***0.69***
	Explain			1.00	•	*0.34*	***0.83***
	Analyze				•	•	•
	Apply					1.00	***0.45***

	Paper 2	**Memorize**	**Procedure**	**Explain**	**Analyze**	**Apply**	**Total mark**
HG (n = 975)	**Memorize**	1.00	*0.64*	*0.92*	*0.79*	*0.82*	***0.97***
	Procedure		1.00	*0.67*	*0.59*	*0.64*	***0.73***
	Explain			1.00	*0.82*	*0.85*	***0.98***
	Analyze				1.00	*0.76*	***0.86***
	Apply					1.00	***0.88***
SG (n = 881)	**Memorize**	1.00	*0.72*	*0.86*	•	*0.69*	***0.98***
	Procedure		1.00	*0.69*	•	*0.59*	***0.77***
	Explain			1.00	•	*0.70*	***0.94***
	Analyze				•	•	•
	Apply					1.00	***0.77***

Table 6.33 Correlations between the same candidates performance in Papers 1 and 2 of the 2006 examinations with respect to cognitive demand.

HG (n = 975)

		PAPER 2					
		Memorize	**Procedure**	**Explain**	**Analyze**	**Apply**	**Total mark**
PAPER 1	**Memorize**	*0.91*	*0.64*	*0.89*	*0.79*	*0.80*	***0.92***
	Procedure	*0.75*	*0.72*	*0.76*	*0.67*	*0.72*	***0.79***
	Explain	*0.85*	*0.65*	*0.86*	*0.76*	*0.77*	***0.88***
	Analyze	*0.62*	*0.52*	*0.59*	*0.53*	*0.53*	***0.62***
	Apply	*0.81*	*0.69*	*0.85*	*0.75*	*0.78*	***0.86***
	Total mark	*0.92*	*0.72*	*0.92*	*0.82*	*0.84*	***0.95***

SG (n=881)

		PAPER 2					
		Memorize	**Procedure**	**Explain**	**Analyze**	**Apply**	**Total mark**
PAPER 1	**Memorize**	*0.90*	*0.72*	*0.85*	•	*0.71*	***0.92***
	Procedure	*0.60*	*0.55*	*0.55*	•	*0.50*	***0.62***
	Explain	*0.73*	*0.63*	*0.69*	•	*0.59*	***0.75***
	Analyze				•	•	•
	Apply	*0.41*	*0.36*	*0.39*		*0.38*	***0.43***
	Total mark	*0.91*	*0.74*	*0.85*		*0.72*	***0.93***

Given that more able students achieved proportionally more marks for each of the PET categories than less able students, the questions "How much more, how much less; or how much HOCS, should be examined?" need to be asked with respect to the performance standards. In addition, "How can cut-scores best be set to separate different categories of student performance, given the suite of student performances presented in this study?"

6.2.2.3 *Performance according to content (topic and cognitive demand)* (Appendices 6.57 & 6.58)

Analyses of categories of students performance (as indicated by their symbols) in content categories may indicate if categories of students have different competencies with respect to content. Student performance in each content category in each of the eight question papers reported as the average percentage mark achieved (Figures 6.20 - 6.23). The reason for using percentages here rather than the actual mark obtained, is that the maximum possible marks obtainable in each content category varied considerably between examinations (Figures 6.24 & 6.25). The use of percentages allows the reader to easily identify which categories of content were passed (i.e., a pass requires obtaining >40% for HG and >33.3% for SG) for each of the performance categories and the relative changes in performance between the performance categories. Figures 6.20 to 6.23 should be read in conjunction with Appendices 6.57 and 6.58 which show when the average student mark was no longer a pass mark for each content category.

2005

A

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	90	**99**	94	92	80
Photosynthesis	95	93	96	**98**	81
Respiration	92	84	82	73	40
Human nutrition	92	**99**	41	82	82
Human gas-eous exchange	89		80	96	
Population dynamics	88	84		96	53

B

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	85	**94**	83	82	72
Photosynthesis	86	89	88	**98**	67
Respiration	81	70	66	55	27
Human nutrition	84	90	21	69	63
Human gas-eous exchange	80		68	92	
Population dynamics	79	85		**95**	41

C

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	75	**94**	74	64	64
Photosynthesis	83	82	72	**88**	57
Respiration	64	55	51	48	16
Human nutrition	73	82	9	55	53
Human gas-eous exchange	68		55	**88**	
Population dynamics	71	78		**92**	33

D

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	64	**91**	62	62	54
Photosynthesis	68	65	56	75	38
Respiration	53	44	39	37	12
Human nutrition	63	79	5	38	46
Human gas-eous exchange	57		44	**88**	
Population dynamics	62	71		**87**	24

E

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	56	**86**	60	61	46
Photosynthesis	59	51	43	65	29
Respiration	37	36	27	34	7
Human nutrition	49	**70**	1	29	37
Human gas-eous exchange	42		28	69	
Population dynamics	57	65		**84**	18

2006

A

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	93				75
Photosynthesis	90	**99**	76	**96**	67
Respiration	95			**96**	
Human nutrition	88	90			83
Human gas-eous exchange	91	83	90		**99**
Population dynamics	85	93	83		85

B

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	80				57
Photosynthesis	77	**97**	62	**93**	43
Respiration	88			**93**	
Human nutrition	73	86			69
Human gas-eous exchange	80	65	78		**97**
Population dynamics	76	73	67		73

C

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	71				41
Photosynthesis	63	**96**	57	**87**	38
Respiration	82			**87**	
Human nutrition	65	81			62
Human gas-eous exchange	68	47	63		**97**
Population dynamics	70	57	59		64

D

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	58				34
Photosynthesis	43	**96**	45	**79**	31
Respiration	65			**79**	
Human nutrition	52	**79**			53
Human gas-eous exchange	55	45	48		**92**
Population dynamics	60	42	53		65

E

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	42				27
Photosynthesis	37	**95**	37	71	22
Respiration	54			71	
Human nutrition	45	**75**			41
Human gas-eous exchange	41	29	31		**87**
Population dynamics	51	29	38		47

Figure 6.20 continued on next page

2005

F

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	42	**82**	36	58	36
Photosynthesis	48	37	34	56	28
Respiration	29	40	21	28	3
Human nutrition	40	**64**	1	17	36
Human gas-eous exchange	33		19	60	
Population dynamics	48	57		**85**	14

FF

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	39	**75**	31	50	29
Photosynthesis	46	30	27	43	17
Respiration	20	34	13	24	2
Human nutrition	34	**69**	1	16	31
Human gas-eous exchange	27		14	60	
Population dynamics	41	58		**75**	14

G

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	29	**67**	24	47	25
Photosynthesis	41	24	23	35	13
Respiration	16	16	12	22	1
Human nutrition	29	52	1	12	24
Human gas-eous exchange	25		14	43	
Population dynamics	37	**56**		**66**	10

GG

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	24	**55**	21	39	21
Photosynthesis	29	19	13	24	13
Respiration	10	18	11	22	1
Human nutrition	24	39	1	8	27
Human gas-eous exchange	21		9	42	
Population dynamics	30	**43**		**62**	4

H

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	14	**43**	8	33	11
Photosynthesis	18	8	6	14	4
Respiration	3	10	8	12	1
Human nutrition	14	30	0	6	14
Human gas-eous exchange	12		7	25	
Population dynamics	19	27		**56**	4

2006

F

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	26				25
Photosynthesis	27	**87**	28	64	12
Respiration	43			64	
Human nutrition	38	**71**			33
Human gas-eous exchange	31	23	22		**84**
Population dynamics	40	20	32		44

FF

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	18				21
Photosynthesis	24	**82**	22	45	14
Respiration	36			45	
Human nutrition	37	**67**			29
Human gas-eous exchange	27	19	16		**75**
Population dynamics	38	18	23		37

G

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	20				17
Photosynthesis	21	**80**	17	43	8
Respiration	38			43	
Human nutrition	33	**60**			20
Human gas-eous exchange	23	13	13		**71**
Population dynamics	25	10	21		33

GG

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	7				17
Photosynthesis	17	**74**	18	26	2
Respiration	26			26	
Human nutrition	30	**52**			19
Human gas-eous exchange	18	11	8		**54**
Population dynamics	23	5	15		22

H

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	5				14
Photosynthesis	13	**61**	8	15	6
Respiration	25			15	
Human nutrition	27	38			19
Human gas-eous exchange	13	4	6		**32**
Population dynamics	15	1	9		15

Figure 6.20 Average percentage obtained in content categories within achievement levels, 2005 and 2006 HG Paper 1. Blank cell = no questions; shaded cell = pass [≥ 40 %]; bold numbers = highest passing percentages.

2005	A					B					C					D					E				
	Memorize	Procedure	Explain	Analyze	Apply	Memorize	Procedure	Explain	Analyze	Apply	Memorize	Procedure	Explain	Analyze	Apply	Memorize	Procedure	Explain	Analyze	Apply	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	93			81	80	82			72	49	71			53	20	63			46	21	53			26	8
Plant water relations	88	**97**	85	90	67	80	**92**	77	82	58	76	**91**	64	77	48	67	**85**	51	66	40	58	**79**	43	57	29
Human excretion	93	91	75	**95**	86	86	**90**	57	**95**	76	78	**86**	47	**90**	72	69	**85**	30	**84**	57	60	**75**	23	**80**	30
Human co-ordination	90		84	74	65	83		78	52	59	75		67	30	46	61		51	17	40	44		39	11	31
Thermo-regulation	**94**		75	81		80		58	68		72		43	50		63		28	39		54		18	36	

2006	A					B					C					D					E				
	Memorize	Procedure	Explain	Analyze	Apply	Memorize	Procedure	Explain	Analyze	Apply	Memorize	Procedure	Explain	Analyze	Apply	Memorize	Procedure	Explain	Analyze	Apply	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	85		90			73		80			**73**		69			64		41			**57**		47		
Plant water relations	**95**	85	86		61	82	**83**	74		51	72	**83**	62		46	60	**74**	46		43	48	**72**	39		35
Human excretion	**96**	85	74		77	**90**	76	63		62	**87**	72	53		41	**78**	63	41		33	**68**	**57**	27		23
Human co-ordination	**91**		83	**91**	66	**84**		72	**84**	55	**73**		62	69	44	**67**		52	59	32	55		39	50	24
Thermo-regulation			81	67	73			72	63	58			56	55	48			45	51	40			34	48	37

Figure 6.21 continued on next page

2005

F

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	35			27	8
Plant water relations	48	**69**	37	45	25
Human excretion	50	**67**	11	**66**	31
Human co-ordination	34		34	7	25
Thermo-regulation	46		10	22	

FF

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	27			20	3
Plant water relations	45	**62**	32	39	22
Human excretion	45	**57**	6	**55**	18
Human co-ordination	26		24	7	23
Thermo-regulation	38		8	18	

G

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	22			11	4
Plant water relations	35	**58**	29	36	18
Human excretion	38	**45**	6	**46**	18
Human co-ordination	21		25	5	20
Thermo-regulation	37		6	14	

GG

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	12			6	2
Plant water relations	31	**50**	27	25	12
Human excretion	28	**41**	2	39	20
Human co-ordination	17		16	3	17
Thermo-regulation	31		3	11	

H

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	7			4	1
Plant water relations	24	21	21	15	9
Human excretion	18	21	1	22	13
Human co-ordination	11		10	2	9
Thermo-regulation	25		2	5	

2006

F

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	49		29		
Plant water relations	37	**64**	30		23
Human excretion	**61**	**51**	20		16
Human co-ordination	48		29	36	15
Thermo-regulation			25	40	25

FF

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	**43**		23		
Plant water relations	30	**67**	25		25
Human excretion	**50**	**43**	16		17
Human co-ordination	41		25	28	16
Thermo-regulation			20	36	23

G

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	33		20		
Plant water relations	28	**63**	21		20
Human excretion	**42**	**42**	8		16
Human co-ordination	36		18	24	11
Thermo-regulation			16	35	17

GG

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	32		7		
Plant water relations	24	**57**	15		15
Human excretion	36	39	6		11
Human co-ordination	30		14	12	7
Thermo-regulation			11	26	10

H

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	22		4		
Plant water relations	13	**42**	10		9
Human excretion	23	20	4		6
Human co-ordination	22		7	6	4
Thermo-regulation			9	17	4

Figure 6.21 Average percentage obtained in content categories within achievement levels, 2005 and 2006 HG Paper 2. Blank cell = no questions; shaded cell = pass [≥ 40 %]; bold numbers = highest passing percentages.

2005

A

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	80	77			**91**
Photosynthesis	90				
Respiration	83		73		
Human nutrition	89		64		
Human gas-eous exchange	88		84	67	
Population dynamics	**92**	**96**	29		87

B

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	66	69			82
Photosynthesis	**83**				
Respiration	70		60		
Human nutrition	79		51		
Human gas-eous exchange	79		68	50	
Population dynamics	**85**	**93**	18		75

C

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	61	63			68
Photosynthesis	**75**				
Respiration	59		51		
Human nutrition	67		39		
Human gas-eous exchange	68		53	41	
Population dynamics	**78**	**90**	18		61

D

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	49	60			51
Photosynthesis	**63**				
Respiration	50		42		
Human nutrition	54		27		
Human gas-eous exchange	57		35	30	
Population dynamics	**75**	**93**	11		55

E

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	45	**56**			42
Photosynthesis	49				
Respiration	40		27		
Human nutrition	45		20		
Human gas-eous exchange	43		20	22	
Population dynamics	69	**83**	7		45

2006

A

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	82	91	77		
Photosynthesis	84		67		
Respiration	91		**98**		
Human nutrition	81				
Human gas-eous exchange	79		50		
Population dynamics	89	**97**	**95**		71

B

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	71	**86**	65		
Photosynthesis	75		64		
Respiration	**86**		81		
Human nutrition	67				
Human gas-eous exchange	71		34		
Population dynamics	**86**	**99**	**93**		53

C

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	58	78	56		
Photosynthesis	61		56		
Respiration	69		**90**		
Human nutrition	56				
Human gas-eous exchange	64		18		
Population dynamics	**80**	**96**	**80**		45

D

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	45	72	47		
Photosynthesis	47		45		
Respiration	56		**84**		
Human nutrition	46				
Human gas-eous exchange	54		10		
Population dynamics	**73**	**90**	72		41

E

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	36	61	39		
Photosynthesis	30		34		
Respiration	38		**80**		
Human nutrition	33				
Human gas-eous exchange	46		5		
Population dynamics	**66**	**88**	65		35

Figure 6.22 continued on next page

2005

F

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	39	**49**			32
Photosynthesis	38				
Respiration	33		25		
Human nutrition	34		13		
Human gaseous exchange	34		16	19	
Population dynamics	**62**	**79**	11		34

FF

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	38	**48**			26
Photosynthesis	32				
Respiration	28		16		
Human nutrition	28		7		
Human gaseous exchange	27		6	15	
Population dynamics	**57**	**77**	6		34

G

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	34	**38**			24
Photosynthesis	27				
Respiration	25		15		
Human nutrition	26		4		
Human gaseous exchange	21		11	12	
Population dynamics	**50**	**72**	3		28

GG

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	30	**38**			20
Photosynthesis	19				
Respiration	19		15		
Human nutrition	20		4		
Human gaseous exchange	18		11	12	
Population dynamics	**45**	**72**	3		28

H

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	26	22			16
Photosynthesis	12				
Respiration	11		4		
Human nutrition	13		1		
Human gaseous exchange	7		3	6	
Population dynamics	26	**47**	2		14

2006

F

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	27	47	30		
Photosynthesis	23		28		
Respiration	30		**73**		
Human nutrition	26				
Human gaseous exchange	40		3		
Population dynamics	**59**	**81**	56		33

FF

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	21	43	26		
Photosynthesis	19		20		
Respiration	24		**72**		
Human nutrition	18				
Human gaseous exchange	35		1		
Population dynamics	**55**	**78**	42		33

G

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	18	38	21		
Photosynthesis	13		14		
Respiration	21		**74**		
Human nutrition	14				
Human gaseous exchange	32		1		
Population dynamics	48	**77**	**50**		33

GG

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	14	27	14		
Photosynthesis	7		12		
Respiration	19		**70**		
Human nutrition	10				
Human gaseous exchange	31		4		
Population dynamics	38	**61**	**42**		26

H

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	11	14	6		
Photosynthesis	5		8		
Respiration	12		**75**		
Human nutrition	7				
Human gaseous exchange	24		0		
Population dynamics	28	**50**	**34**		23

Figure 6.22 Average percentage obtained in content categories within achievement levels, 2005 and 2006 SG Paper 1. Blank cell = no questions; shaded cell = pass [≥ 33.3 %]; bold numbers = highest passing percentages.

2005

A

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	**100**		**100**		
Plant water relations	88	98	81		70
Human excretion	86		69		58
Human co-ordination	91		93		67
Thermo-regulation	86	61	**100**		

B

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	**86**		**94**		
Plant water relations	80	84	74		55
Human excretion	79		43		43
Human co-ordination	80		82		55
Thermo-regulation	73	57	**93**		

C

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	**90**		**95**		
Plant water relations	75	74	54		50
Human excretion	69		24		33
Human co-ordination	70		69		50
Thermo-regulation	57	51	**81**		

D

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	**81**		**89**		
Plant water relations	63	63	43		39
Human excretion	62		14		22
Human co-ordination	60		55		44
Thermo-regulation	46	49	**70**		

E

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	**67**		**86**		
Plant water relations	50	56	31		34
Human excretion	49		5		18
Human co-ordination	48		57		44
Thermo-regulation	38	44	**62**		

2006

A

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	76				
Plant water relations	**97**	**94**	73		92
Human excretion	81		64		90
Human co-ordination	**93**	75	73		73
Thermo-regulation	67		75		71

B

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	73				
Plant water relations	**86**	74	64		63
Human excretion	70		67		71
Human co-ordination	**82**	**80**	59		66
Thermo-regulation	61		70		58

C

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	68				
Plant water relations	**81**	60	49		57
Human excretion	67		50		63
Human co-ordination	**75**	**75**	40		62
Thermo-regulation	46		60		42

D

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	55				
Plant water relations	**72**	54	34		38
Human excretion	57		46		60
Human co-ordination	**65**	**68**	27		52
Thermo-regulation	34		48		40

E

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	51				
Plant water relations	**62**	32	27		24
Human excretion	49		38		51
Human co-ordination	**52**	**58**	17		53
Thermo-regulation	30		35		32

Figure 6.23 continued on next page

2005

F

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	**62**		**84**		
Plant water relations	43	43	23		26
Human excretion	42		3		17
Human co-ordination	37		41		38
Thermo-regulation	29	42	**58**		

FF

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	**57**		**84**		
Plant water relations	34	38	20		24
Human excretion	33		1		16
Human co-ordination	33		38		35
Thermo-regulation	24	37	**52**		

G

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	**54**		**77**		
Plant water relations	30	32	17		19
Human excretion	30		2		14
Human co-ordination	28		41		33
Thermo-regulation	20	35	**46**		

GG

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	**45**		**80**		
Plant water relations	23	24	15		16
Human excretion	24		1		15
Human co-ordination	22		27		28
Thermo-regulation	19	30	**34**		

H

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	33		**61**		
Plant water relations	12	9	9		6
Human excretion	15		0		12
Human co-ordination	14		24		32
Thermo-regulation	9	21	18		

2006

F

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	36				
Plant water relations	**49**	28	16		16
Human excretion	42		36		44
Human co-ordination	**45**	**49**	9		**47**
Thermo-regulation	18		26		25

FF

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	30				
Plant water relations	**40**	21	12		11
Human excretion	**40**		31		33
Human co-ordination	**40**	**42**	9		**44**
Thermo-regulation	13		13		23

G

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	23				
Plant water relations	33	16	8		8
Human excretion	35		26		**36**
Human co-ordination	**36**	**39**	6		**45**
Thermo-regulation	9		14		20

GG

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	19				
Plant water relations	25	17	6		9
Human excretion	29		14		30
Human co-ordination	30	**34**	3		**43**
Thermo-regulation	7		7		19

H

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	11				
Plant water relations	14	11	3		5
Human excretion	**18**		10		14
Human co-ordination	**20**	**18**	2		**32**
Thermo-regulation	3		3		15

Figure 6.23 Average percentage obtained in content categories within achievement levels, 2005 and 2006 SG Paper 2. Blank cell = no questions; shaded cell = pass (≥ 33.3 %); bold numbers = highest passing percentages.

A. **2005 HG**

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	11	3	2	1	18
Photosynthesis	10	8	3	3	3
Respiration	5	2	6	8	7
Human nutrition	22	1	4	8	5
Human gaseous exchange	27		3	2	
Population dynamics	21	5		1	8

2006 SG

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	9				20
Photosynthesis	17	2	4	3.5	2
Respiration	8			3.5	
Human nutrition	20	19			8
Human gaseous exchange	30	9	7		2
Population dynamics	5	6	18		7

B. **2005 SG**

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	13	6			6
Photosynthesis	22				
Respiration	12		2		
Human nutrition	26		9		
Human gaseous exchange	22		2	3	
Population dynamics	8	6	4		9

2006 SG

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	25	6	10		
Photosynthesis	15		4		
Respiration	15		5		
Human nutrition	21				
Human gaseous exchange	20		2		
Population dynamics	19	3	2		3

Figure 6.24 Total possible marks for content categories – Paper 1. Blank cells indicate no questions.

A. **2005 HG**

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	3			2	3
Plant water relations	13	13	9	8	17
Human excretion	16	8	10	3	2
Human co-ordination	31		7	13	17
Thermo-regulation	6		7	7	

2006 HG

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	8		2		
Plant water relations	26	4	18		8
Human excretion	3	12	18		5
Human co-ordination	32		20	8	7
Thermo-regulation			13	7	3

B. **2005 SG**

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	4		2		
Plant water relations	20	10	8		9
Human excretion	17		6		10
Human co-ordination	40		5		2
Thermo-regulation	11	3	3		

2006 SG

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	7				
Plant water relations	16	3	19		2
Human excretion	21		7		5
Human co-ordination	33	6	11		5
Thermo-regulation	3		6		4

Figure 6.25 Total possible marks for content categories – Paper 2. Blank cells indicate no questions.

Student performance with respect to content for each comparable question paper (e.g., 2005 HG Paper 1 and 2006 HG Paper 1) differed between the two years because the content standards and the structures of the question papers varied between the years (Sections 6.1.1 & 6.1.2). The pattern in the HG question papers was for the Apply questions to be failed first, suggesting that these questions were the most difficult for students. The same pattern was not obvious in the SG question paper where there were fewer Analyze and Apply questions and many of the questions requiring the application of knowledge were questions that did not require candidates to move through the Analyze category as defined by PET in Chapter 4 (Chapter 4, Figure 4.7 [ii]). Some content categories did not discriminate between students who achieved different symbols were: Biochemistry and Perform-Routine-Procedures, Population dynamics and Analyze (2005 HG Paper 1); Photosynthesis and Perform-Routine-Procedures (2006 HG Paper 1); Water relations and Perform-Routine-Procedures (2006 HG Paper 2); Population dynamics and Perform-Routine-Procedures (2005 SG Paper 1), Plant hormones and Explain (2005 SG Paper 2); and Population dynamics and Perform-Routine-Procedures, Population dynamics and Explain, Respiration and Explain (2006 SG Paper 1). None of the non-discriminating content categories totalled more than six marks and they thus contributed only a small part to the total score. The unequal mark allocation between the content categories means that it is difficult to determine if students would have performed differently had they been afforded equal opportunity to demonstrate competence in each content category and if questions in each content category had the same question structure and the same expected answer structure.

Viewing student performance from the perspective of content does bring particular insights for trying to understand the total marks that different students achieve for a question paper, especially because topics provide the context for HOCS. For example, most students are able to demonstrate their ability to Perform-Routine-Procedures irrespective of topic. Post-2000, Performing-Routine-Procedures comprised relatively larger percentages of the HG question papers that in previous years (Tables 6.18 & 6.19). As a result, one might expect that teachers emphasized the teaching of routine procedures more, in order that students achieved better marks in corresponding questions.

Conceptualizing content as two-dimensional (i.e., topic and performance expectation) is in line with the principles of the CBS which required through the objectives and approaches to the syllabus that all levels of cognitive demand be infused through all the topics (called content in the CBS) taught (Chapter 2, Figure 2.4; Chapter 4, Tables 4.3 & 4.4). In order to make productive use of knowledge of the different topics, such as in application of knowledge, students need to be able to retain the knowledge, and this requires that students are able to make sense of, and ascribe meaning, to the knowledge (Sousa, 1996). Whether SC students were taught to make sense or meaning of the facts, concepts and processed in the enacted curriculum, is questioned, and possibly unknowable, given that lack of opportunity to demonstrate HOCS in the SC examinations.

6.2.3 Reliability of student performance data

Student performance has been discussed by providing descriptions of the performance standards across the continuum according to the structural characteristics of the question paper and the content standards. In addition, relationships between the different aspects of content standards, that is, topics and performance expectations, and student performance have been examined. However, the educational consequentiality of these results depends on the reliability of the student marks used to measure these different constructs. The student marks for each of these different constructs was obtained by adding the marks obtained in all the scorable events which comprised each construct.

This section describes reliability analyses that were undertaken to explore the internal consistency of different ways in which that the student performance data has been grouped in this study. These groups are, groups of individual scorable event marks; or as composite marks comprising, topics, performance expectations or content categories The co-efficient of reliability used here was Cronbach's alpha of which values of 0.70 are generally considered to be acceptable as an indication of internal consistency (McMillan & Schumacher, 2001). Exploratory factor analyses were then performed to test the presence of underlying structures and, if present, their dimensionality, in the student performance data. For certification purposes SC Biology students are classified according to only the total marks achieved (i.e., the sum of the marks for each of the scorable events). Therefore, discriminant functions analyses were used to test the ability of the vector combination of marks achieved in individual scorable events to confirm, or to not confirm, the aggregate classifications.

6.2.3.1 *Internal reliability*

Most of the Cronbach's alpha values for different combinations of scorable according to the different topics and some categories of performance expectations were greater than 0.70, indicating a conventional internal consistency in these groupings of the data (Tables 6.34 - 6.41). However, the low average inter-item correlations between the items comprising each of these constructs suggests a great deal of, a that is as yet unexplained, variation in student performance. The high Cronbach's alpha values could result of the large number of individual items in some of the groupings.

Table 6.34 Results of reliability analyses of groupings of scorable events: 2005 HG Paper 1.

Grouping	Variable(s)	No. items	Cronbach's alpha	Average inter-item correlation
All	Individual scorable events	81	.96	.31
Topics	Biochemistry	16	.89	.35
	Photosynthesis	11	.81	.36
	Respiration	12	.84	.34
	Human nutrition	19	.89	.34
	Human gaseous exchange	8	.66	.39
	Population dynamics	14	.80	.23
Topics	Biochemistry, Photosynthesis, Respiration, Human nutrition, Human gaseous exchange, Population dynamics	6	.96	.80
Cognitive demand	Memorize	40	.92	.33
	Procedure	8	.63	.26
	Explain	7	.79	.41
	Analyze	11	.77	.25
	Apply	14	.86	.32
Cognitive demand	Memorize, Perform-Routine-Procedures, Explain, Analyze, Apply	5	.72	.81
Content	Biochemistry & Memorize	5	.75	.42
	Biochemistry & Perform-Routine-Procedures	2	.24	.19
	Biochemistry & Explain	1	—	—
	Biochemistry & Analyze	1	—	—
	Biochemistry & Apply	7	.79	.37
	Photosynthesis & Memorize	5	.63	.31
	Photosynthesis & Perform-Routine-Procedures	2	.32	.37
	Photosynthesis & Explain	2	.54	.44
	Photosynthesis & Analyze	1	—	—
	Photosynthesis & Apply	1	—	—
	Respiration & Memorize	3	.67	.42
	Respiration & Perform-Routine-Procedures	1	—	—
	Respiration & Explain	2	.44	.42
	Respiration & Analyze	4	.58	.27
	Respiration & Apply	2	.65	.49
	Human nutrition & Memorize	12	.86	.37
	Human nutrition & Perform-Routine-Procedures	1	—	—
	Human nutrition & Explain	1	—	—
	Human nutrition & Analyze	3	.56	.36
	Human nutrition & Apply	2	.71	.58
	Human gaseous exchange & Memorize	6	.54	.76
	Human gaseous exchange & Explain	1	—	—
	Human gaseous exchange & Analyze	1	—	—
	Population dynamics & Memorize	9	.74	.27
	Population dynamics & Perform-Routine-Procedures	2	.33	.21
	Population dynamics & Analyze	1	—	—
	Population dynamics & Apply	2	.29	.20
Content	Biochemistry & Memorize to Population dynamics & Apply	27	.93	.45

Note: A ó– ómeans there were no questions in that content category

Table 6.35 Results of reliability analyses of groupings of scorable events: 2005 HG Paper 2.

Grouping	Variable(s)	No. items	Cronbach's alpha	Average inter-item correlation
All	Individual scorable events	78	.95	.30
Topics	Plant hormones	5	.75	.43
	Plant water relationships	24	.86	.27
	Human excretion	14	.85	.36
	Human co-ordination	27	.85	.32
	Thermoregulation	8	.80	.38
Topics	Plant hormones, Plant water relations, Human excretion, Human co-ordination, Thermoregulation	5	.88	.81
Cognitive demand	Memorize	33	.85	.28
	Procedure	3	.38	.31
	Explain	10	.87	.43
	Analyze	15	.87	.36
	Apply	17	.84	.27
Cognitive demand	Memorize, Perform-Routine-Procedures, Explain, Analyze, Apply	5	.90	.81
Content	Plant hormones & Memorize	2	.55	.44
	Plant hormones & Analyze	1	—	—
	Plant hormones & Apply	2	.76	.70
	Plant water relations & Memorize	9	.64	.22
	Plant water relations & Perform-Routine-Procedures	2	.45	.57
	Plant water relations & Explain	3	.51	.28
	Plant water relations & Analyze	4	.60	.33
	Plant water relations & Apply	6	.64	.30
	Human excretion & Memorize	9	.78	.34
	Human excretion & Perform-Routine-Procedures	1	—	—
	Human excretion & Explain	2	.69	.62
	Human excretion & Analyze	1	—	—
	Human excretion & Apply	1	—	—
	Human co-ordination & Memorize	9	.58	.31
	Human co-ordination & Explain	3	.65	.39
	Human co-ordination & Analyze	7	.79	.47
	Human co-ordination & Apply	8	.75	.28
	Thermoregulation & Memorize	4	.49	.26
	Thermoregulation & Explain	2	.69	.56
	Thermoregulation & Analyze	2	.61	.47
Content	Plant hormones & Memorize to Thermoregulation & Apply	20	.93	.55

Note: A ó–ómeans there were no questions in that content category

Table 6.36 Results of reliability analyses of groupings of scorable events: 2005 SG Paper 1.

Grouping	Variable(s)	No. items	Cronbach's alpha	Average inter-item correlation
All	Individual scorable events	67	.96	.29
Topics	Biochemistry	11	.68	.21
	Photosynthesis	9	.84	.40
	Respiration	8	.72	.24
	Human nutrition	16	.88	.33
	Human gaseous exchange	12	.86	.39
	Population dynamics	11	.73	.21
Topics	Biochemistry, Photosynthesis, Respiration, Human nutrition, Human gaseous exchange, Population dynamics	6	.93	.74
Cognitive demand	Memorize	51	.95	.33
	Procedure	4	.38	.18
	Explain	6	.75	.36
	Analyze	1	—	—
	Apply	5	.70	.33
Cognitive demand	Memorize, Perform-Routine-Procedures, Explain, Analyze, Apply	5	.51	.56
Content	Biochemistry & Memorize	8	.55	.17
	Biochemistry & Perform-Routine-Procedures	1	—	—
	Biochemistry & Apply	2	.70	.54
	Photosynthesis & Memorize	9	.84	.40
	Respiration & Memorize	7	.68	.24
	Respiration & Explain	1	—	—
	Human nutrition & Memorize	13	.83	.36
	Human nutrition & Explain	3	.63	.45
	Human gaseous exchange & Memorize	10	.83	.40
	Human gaseous exchange & Explain	1	—	—
	Human gaseous exchange & Analyze	1	—	—
	Population dynamics & Memorize	4	.53	.23
	Population dynamics & Perform-Routine-Procedures	3	.35	.17
	Population dynamic & Explain	1	—	—
	Population dynamics & Apply	3	.51	.28
Content categories	Biochemistry & Memorize to Population dynamics & Apply	15	.90	.48

Note: A ó— ómeans there were no questions in that content category

Table 6.37 Results of reliability analyses of groupings of scorable events: 2005 SG Paper 2.

Grouping	Variable(s)	No. items	Cronbach's alpha	Average inter-item correlation
All	Individual scorable events	67	.94	.21
Topics	Plant hormones	3	.29	.13
	Plant water relationships	19	.81	.22
	Human excretion	16	.81	.26
	Human co-ordination	21	.85	.22
	Thermoregulation	8	.65	.19
Topics	Plant hormones, Plant water relations, Human excretion, Human co-ordination, Thermoregulation	5	.84	.65
Cognitive demand	Memorize	45	.92	.24
	Procedure	4	.29	.16
	Explain	10	.63	.16
	Apply	8	.61	.16
Cognitive demand	Memorize, Perform-Routine-Procedures, Explain, Apply	4	.59	.62
Content	Plant hormones & Memorize	2	.20	.11
	Plant hormones & Explain	1	—	—
	Plant water relations & Memorize	9	.75	.27
	Plant water relations & Perform-Routine-Procedures	2	.24	.23
	Plant water relations & Explain	4	.227	.10
	Plant water relations & Apply	4	.54	.23
	Human excretion & Memorize	11	.77	.27
	Human excretion & Explain	2	.57	.42
	Human excretion & Apply	3	.44	.21
	Human co-ordination & Memorize	18	.84	.24
	Human co-ordination & Explain	2	.33	.21
	Human co-ordination & Apply	1	—	—
	Thermoregulation & Memorize	5	.58	.23
	Thermoregulation & Perform-Routine-Procedures	2	.02	.02
	Thermoregulation & Explain	1	—	—
Content	Plant hormones & Memorize to Thermoregulation & Apply	15	.85	.36

Note: A ó—ómeans there were no questions in that content category

Table 6.38 Results of reliability analyses of groupings of scorable events: 2006 HG Paper 1.

Grouping	Variable(s)	No. items	Cronbach's alpha	Average inter-item correlation
All	Individual scorable events	68	.94	.27
Topics	Biochemistry	11	.76	.34
	Photosynthesis	8	.78	.34
	Respiration	6	.62	.32
	Human nutrition	15	.72	.21
	Human gaseous exchange	17	.77	.29
	Population dynamics	11	.80	.32
Topics	Biochemistry, Photosynthesis, Respiration, Human nutrition, Human gaseous exchange, Population dynamics	6	.92	.77
Cognitive demand	Memorize	30	.87	.27
	Procedure	9	.70	.32
	Explain	9	.82	.39
	Analyze	2	1.00	1.00
	Apply	18	.85	.26
Cognitive demand	Memorize, Perform-Routine-Procedures, Explain, Analyze, Apply	5	.83	.72
Content	Biochemistry & Memorize	2	.29	.56
	Biochemistry & Apply	9	.78	.32
	Photosynthesis & Memorize	3	.58	.35
	Photosynthesis & Perform-Routine-Procedures	1	–	–
	Photosynthesis & Explain	2	.54	.37
	Photosynthesis & Analyze	1	–	–
	Photosynthesis & Apply	1	–	–
	Respiration & Memorize	5	.61	.29
	Respiration & Analyze	1	–	–
	Human nutrition & Memorize	8	.65	.23
	Human nutrition & Perform-Routine-Procedures	3	.25	.26
	Human nutrition & Apply	2	.59	.26
	Human gaseous exchange & Memorize	9	.53	.22
	Human gaseous exchange & Perform-Routine-Procedures	4	.72	.41
	Human gaseous exchange & Explain	3	.72	.46
	Human gaseous exchange & Apply	1	–	–
	Population dynamics & Memorize	3	.48	.31
	Population dynamics & Perform-Routine-Procedures	1	–	–
	Population dynamics & Explain	4	.63	.36
	Population dynamics & Apply	3	.47	.24
Content	Biochemistry & Memorize to Population dynamics & Apply	20	.93	.52

Table 6.39 Results of reliability analyses of groupings of scorable events: 2006 HG Paper 2.

Grouping	Variable(s)	No. items	Cronbach's alpha	Average inter-item correlation
All	Individual scorable events	75	.96	.30
Topics	Plant hormones	6	.62	.26
	Plant water relationships	22	.83	.29
	Human excretion	10	.80	.37
	Human co-ordination	27	.91	.31
	Thermoregulation	10	.84	.34
Topics	Plant hormones, Plant water relations, Human excretion, Human co-ordination, Thermoregulation	5	.89	.79
Cognitive demand	Memorize	32	.87	.27
	Procedure	5	.15	.20
	Explain	24	.93	.40
	Analyze	6	.57	.29
	Apply	8	.79	.32
Cognitive demand	Memorize, Perform-Routine-Procedures, Explain, Analyze, Apply	5	.84	.77
Content	Plant hormones & Memorize	5	.50	.21
	Plant hormones & Explain	1	—	—
	Plant water relations & Memorize	9	.56	.30
	Plant water relations & Perform-Routine-Procedures	3	.18	.13
	Plant water relations & Explain	7	.81	.40
	Plant water relations & Apply	3	.60	.34
	Human excretion & Memorize	2	.44	.33
	Human excretion & Perform-Routine-Procedures	2	.53	.45
	Human excretion & Explain	4	.68	.44
	Human excretion & Apply	2	.47	.31
	Human co-ordination & Memorize	16	.83	.27
	Human co-ordination & Explain	7	.80	.39
	Human co-ordination & Analyze	2	.21	.30
	Human co-ordination & Apply	2	.61	.45
	Thermoregulation & Explain	5	.80	.46
	Thermoregulation & Analyze	4	.52	.22
Content	Plant hormones & Memorize to Thermoregulation & Apply	17	.93	.56

Note: A ó— ómeans there were no questions in that content category

Table 6.40 Results of reliability analyses of groupings of scorable events: 2006 SG Paper 1.

Grouping	Variable(s)	No. items	Cronbach's alpha	Average inter-item correlation
All	Individual scorable events	58	.94	.24
Topics	Biochemistry	16	.86	.31
	Photosynthesis	7	.70	.31
	Respiration	10	.70	.22
	Human nutrition	5	.72	.41
	Human gaseous exchange	10	.62	.64
	Population dynamics	10	.74	.23
Topics	Biochemistry, Photosynthesis, Respiration, Human nutrition, Human gaseous exchange, Population dynamics	6	.92	.71
Cognitive demand	Memorize	45	.93	.26
	Procedure	4	.57	.24
	Explain	8	.62	.19
	Apply	1	—	—
Cognitive demand	Memorize, Perform-Routine-Procedures, Explain, Apply	4	.40	.54
Content	Biochemistry & Memorize	10	.84	.36
	Biochemistry & Perform-Routine-Procedures	3	.52	.27
	Biochemistry & Explain	3	.44	.29
	Photosynthesis & Memorize	6	.64	.31
	Photosynthesis & Explain	1	—	—
	Respiration & Memorize	8	.70	.30
	Respiration & Explain	2	.07	.07
	Human nutrition & Memorize	5	.71	.41
	Human gaseous exchange & Memorize	9	.58	.14
	Human gaseous exchange & Explain	1	—	—
	Population dynamics & Memorize	7	.66	.23
	Population dynamics & Perform-Routine-Procedures	1	—	—
	Population dynamics & Explain	1	—	—
	Population dynamics & Apply	1	—	—
Content	Biochemistry & Memorize to Population dynamics & Apply	14	.89	.42

Note: A ó— ómeans there were no questions in that content category

Table 6.41 Results of reliability analyses of groupings of scorable events: 2006 SG Paper 2.

Grouping	Variable(s)	No. items	Cronbach's alpha	Average inter-item correlation
All	Individual scorable events	61	.94	.25
Topics	Plant hormones	4	.54	.27
	Plant water relationships	16	.88	.36
	Human excretion	11	.74	.24
	Human co-ordination	23	.86	.22
	Thermoregulation	7	.78	.34
Topics	Plant hormones, Plant water relations, Human excretion, Human co-ordination, Thermoregulation	5	.85	.70
Cognitive demand	Memorize	39	.91	.24
	Procedure	2	.46	.37
	Explain	13	.86	.38
	Apply	7	.58	.18
Cognitive demand	Memorize, Perform-Routine-Procedures, Explain, Apply	4	.73	.72
Content	Plant hormones & Memorize	4	.54	.27
	Plant water relations & Memorize	8	.79	.36
	Plant water relations & Perform-Routine-Procedures	1	–	–
	Plant water relations & Explain	6	.74	.37
	Plant water relations & Apply	1	–	–
	Human excretion & Memorize	9	.64	.22
	Human excretion & Explain	1	–	–
	Human excretion & Apply	1	–	–
	Human co-ordination & Memorize	16	.79	.21
	Human co-ordination & Perform-Routine-Procedures	1	–	–
	Human co-ordination & Explain	3	.61	.38
	Human co-ordination & Apply	3	.38	.20
	Thermoregulation & Memorize	2	.35	.21
	Thermoregulation & Explain	3	.78	.54
	Thermoregulation & Apply	2	.61	.48
Content	Plant hormones & Memorize to Thermoregulation & Apply	15	.89	.49

Note: A ó– ómeans there were no questions in that content category

Where Cronbach's alpha values were less than 0.70, they were always associated with the PET category Perform-Routine-Procedures which collectively described a group of fairly heterogeneous routines which students might be required to perform (Chapter 4, Figure 4.5), and each of which required completely different suites of particular skills. It might therefore be expected that different students had quite unequal competencies in various combinations of these skills, hence accounting for the lower Cronbach's alpha values for the corresponding scores.

The lack of internal consistency in the data comprising the various content categories was possibly due to the small number of items in each of the categories.

6.2.3.2 *Factor analyses*

Exploratory factor analyses were used to identify and confirm the possible existence of functional topic and performance expectations groupings of the scorable events. Given that the SC Biology examinations were written to address different topics and different levels of cognitive demand, it was expected that distinct factors identifying the constructs, topics and cognitive demand, might have emerged in factor analyses of the student performance data.

The student performance data were combined into five different data sets, representing different combinations of the performance data, for each of the eight question papers, and each was subjected to factor analysis. The five different data sets were the marks achieved for individual scorable events; each of the topics *and* each of the performance expectations; each of the topics; each of the performance expectations; and each of the content categories. Each of the factor analyses was conducted without limiting the number of factors to be loaded.

The factor analyses solutions indicated the percentages of the variance explained and the cumulative variances of the factors which emerged on each of the analyses (Tables 6.42 - 6.46). Despite the fact that three factors consistently emerged for each of the eight factor analyses which used the marks for individual scorable events in a question paper, these three factors counted for at most about 38% of the total cumulative variance (Table 6.42). Only one factor consistently loaded, explaining as much as 84% when the marks obtained in each topic and each category of performance expectations were used (Table 6.43). Similarly, only one factor consistently loaded, explaining as much as 85% when the marks of each topic or each category of performance expectations were used alone (Tables 6.44 & 6.45). Between one and three factors loaded when categories of content were analyzed for potential factors, but considerable variance remained unaccounted for in all these analyses (Table 6.46). When more than one

Table 6.42 Results of factor analyses of 2005 and 2006 script data as individual scorable events.

Year	Question paper	No. scripts	No. scorable events	Factor	Eigenvalue	Total Variance (%)	Cumulative Eigenvalue	Cumulative Variance (%)
2005	**HG Paper 1**	1001	81	1	26.10	32.22	26.10	32.22
				2	2.21	2.728	28.31	34.94
				3	1.63	2.01	29.93	36.95
	HG Paper 2	1004	79	1	25.55	32.76	25.55	32.76
				2	2.65	3.39	28.20	36.15
				3	1.91	2.46	30.11	38.61
	SG Paper 1	982	66	1	20.64	31.27	20.64	31.27
				2	1.78	2.69	22.41	33.96
				3	1.56	2.36	23.97	36.32
	SG Paper 2	854	70	1	16.16	23.09	16.16	23.09
				2	2.25	3.21	18.41	26.30
				3	1.53	2.19	19.95	28.49
2006	**HG Paper 1**	975	67	1	20.38	30.42	20.38	30.42
				2	2.19	3.26	22.57	33.68
				3	1.65	2.46	24.21	36.14
	HG Paper 2	975	78	1	24.13	31.34	24.13	31.34
				2	1.61	2.09	25.74	33.43
				3	1.53	1.98	27.27	35.42
	SG Paper 1	881	58	1	16.08	27.73	16.08	27.73
				2	1.63	2.81	17.71	30.54
				3	1.50	2.59	19.22	33.13
	SG Paper 2	881	62	1	17.41	28.08	17.40	28.08
				2	1.84	2.96	19.24	31.04
				3	1.52	2.44	20.76	33.48

Note: A ó— ómeans there were no questions in that content category

Table 6.43 Results of factor analyses of 2005 and 2006 script data combined into topics and performance expectations.

Year	Question paper	No. scripts	No. topics & performance expectations	Factor	Eigenvalue	Total Variance (%)	Cumulative Eigenvalue	Cumulative Variance (%)
2005	HG Paper 1	1001	11	1	9.19	83.57	9.19	83.57
	HG Paper 2	1004	10	1	8.36	83.59	8.36	83.59
	SG Paper 1	982	11	1	7.75	70.49	7.75	70.49
	SG Paper 2	854	9	1	6.27	69.68	6.27	69.68
2006	HG Paper 1	975	11	1	8.66	78.68	8.66	78.68
	HG Paper 2	975	10	1	8.07	80.65	8.07	80.65
	SG Paper 1	881	10	1	6.95	69.49	6.95	69.49
	SG Paper 2	881	9	1	6.81	75.65	6.81	75.65

Table 6.44 Results of factor analyses of 2005 and 2006 script data combined into topics.

Year	Question paper	No. scripts	No. topics	Factor	Eigenvalue	Total Variance (%)	Cumulative Eigenvalue	Cumulative Variance (%)
2005	HG Paper 1	1001	6	1	4.99	83.13	4.99	83.13
	HG Paper 2	1004	5	1	4.21	84.12	4.21	84.12
	SG Paper 1	982	6	1	4.68	78.08	4.68	78.08
	SG Paper 2	854	5	1	3.55	70.96	3.55	70.96
2006	HG Paper 1	975	6	1	4.86	80.92	4.86	80.92
	HG Paper 2	975	5	1	4.11	82.11	4.11	82.11
	SG Paper 1	881	6	1	4.55	75.91	4.55	75.91
	SG Paper 2	881	5	1	3.77	75.37	3.77	75.37

Table 6.45 Results of factor analyses of 2005 and 2006 script data combined into performance expectations.

Year	Question paper	No. scripts	No. perform-ance expect-ations	Factor	Eigenvalue	Total Variance (%)	Cumulative Eigenvalue	Cumulative Variance (%)
2005	HG Paper 1	1001	5	1	4.24	84.72	4.24	84.72
	HG Paper 2	1004	5	1	4.18	83.71	4.18	83.71
	SG Paper 1	982	5	1	3.19	63.79	3.19	63.79
	SG Paper 2	854	4	1	2.83	70.85	2.83	70.85
2006	HG Paper 1	975	5	1	3.84	76.82	3.84	76.82
	HG Paper 2	975	5	1	4.02	80.38	4.02	80.38
	SG Paper 1	881	4	1	2.58	64.58	2.58	64.58
	SG Paper 2	881	4	1	3.1.3	78.22	3.1.3	78.22

Table 6.46 Results of factor analyses of 2005 and 2006 script data combined into content categories.

Year	Question paper	No. scripts	No. content categor-ies	Factor	Eigenvalue	Total Variance (%)	Cumulative Eigenvalue	Cumulative Variance (%)
2005	HG Paper 1	1001	27	1	12.90	47.79	12.90	47.79
				2	1.09	4.05	13.99	51.83
				3	1.04	3.85	15.03	55.68
	HG Paper 2	1004	20	1	11.56	57.78	11.56	57.79
	SG Paper 1	982	15	1	7.72	51.49	7.72	52.49
	SG Paper 2	854	15	1	6.32	42.11	6.32	42.11
				2	1.04	6.93	7.36	47.04
2006	HG Paper 1	975	20	1	10.54	52.70	10.54	52.70
				2	1.17	5.85	11.71	58.56
	HG Paper 2	975	17	1	9.98	58.72	9.98	58.72
	SG Paper 1	881	14	1	6.61	47.42	6.61	47.42
	SG Paper 2	881	15	1	7.43	53.04	7.43	53.04

factor emerged, inspection of the items which loaded onto each of the factors yielded no rational explanation or interpretation for these factors. The results of the factor analyses suggest that the student performance data represents a unidimensional concept, that is, proficiency, competence or mastery.

The relatively high inter-topic, inter-performance expectations category and inter-paper correlations also supported the existence of a unidimensional, single construct (Tables 6.29 - 6.33). The higher levels of variation explained in the factor analyses of sub-tests of the data according to topics and performance expectations suggested that these groupings are useful ways of organizing student performance data. Cizek, Webb and Kalohn (1995) observed similar unidimensional constructs of mastery in analyses of licensure and certification tests in a health sciences field. Despite the fact that the tests they used had been purposefully written to assess specific levels of HOCS, they failed to provide evidence that the tests could serve this purpose. Cizek et al. (1995) recommended the use of rationally derived categories, like different levels of cognitive demand and topics, even if they might be non-functioning (as determined by factor analyses) in tests, because their use often resulted in more careful test setting and a better balance between the breadth and depth dimensions of test, and therefore improved the validity associated with tests.

6.2.3.3 *Classification of students by aggregate*

Evidence presented in Sections 6.2.3.1 and 6.2.3.2 above suggested that performance standards in all eight SC Biology examinations represent a single, unidimensional construct. In the SC Biology examinations, students were classified as having different levels of mastery according to their aggregate mark in the examinations. The question is therefore: would the individual scorable events or groupings of these scorable events into topics, performance expectations or content categories confirm the classifications made on the aggregate only? Discriminant functions analyses showed that many of the aggregate classifications were confirmed in all 32 analyses (Tables 6.47 - 6.54). Inspection of differing classifications of students indicated that most, but not all, of these were students who had aggregates which were close to the cut-scores of that symbol. As each student classification has implications for the future of a student, these differences are important, especially as a number occur at the highest levels of mastery (i.e., symbols A and B) for both HG and SG, at the interface between pass and fail (E and F – HG; F and FF – SG,) and in the region of HG failures which were converted to SG passes (F, FF, G – HG). This concern is compounded given that so many students in the population represented by the sample in this study performed at these symbols (Chapter 5, Tables 5.1 & 5.2), and that cut-scores were the same within and between years and the same for all different subjects, by fiat

Table 6.47 Classification matrix showing the number of scripts that were correctly and incorrectly classified – 2005 HG Paper 1. Counts of correct classifications on the diagonal.

A. All scorable events — no. variables = 63, WilksôLambda = .0023 , F = 14.34, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	95.00	95	5								
B	93.06	5	94	2							
C	95.00		2	95	3						
D	97.03			3	98						
E	92.86				2	91	5				
F	88.89					2	88	9			
FF	85.86						7	85	7		
G	82.35						1	10	84	7	
GG	86.73								12	85	1
H	85.29									15	87
Total	90.20	100	101	100	103	93	101	104	103	107	88

Note: Scorable events for each of Questions 1.1 to 1.4 combined

B. Topics — no. variables = 6, WilksôLambda = 0.0109, F = 132.89, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	99.00	99	1								
B	93.07	5	94	2							
C	99.00		1	99							
D	97.03			1	98	2					
E	88.78				3	87	8				
F	86.87						86	13			
FF	97.00							97	3		
G	84.31							10	86	6	
GG	93.88								6	92	
H	82.35									18	84
Total	92.11	104	96	102	101	89	94	120	95	116	84

C. Performance expectations — no. variables = 5, WilksôLambda = 0.0112, F = 169.58, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	100									
B	95.05	4	96	1							
C	98.00		2	98							
D	98.02			1	99	1					
E	89.80				2	88	8				
F	87.88						87	12			
FF	95.00							95	5		
G	92.16							6	94	2	
GG	96.94								3	95	
H	84.31									16	86
Total	93.71	104	98	100	101	89	95	113	102	113	86

D. Content — no. variables = 27, WilksôLambda =0.0060, F = 27.93,p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	96.00	96	4								
B	90.10	7	91	3							
C	96.00		3	96	1						
D	99.01				100	1					
E	90.82				2	89	7				
F	85.86					1	85	13			
FF	91.00						3	91	6		
G	84.31							10	86	6	
GG	89.80								10	88	0
H	84.31									16	86
Total	90.71	103	98	99	103	91	95	114	102	110	86

Table 6.48 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2005 HG Paper 2. Counts of correct classifications on the diagonal.

A. All scorable events — no. variables =59, WilksôLambda = .0017, F = 16.52, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	95.10	97	5								
B	92.93	5	92	2							
C	95.05		2	96	3						
D	97.98			1	97	1					
E	95.00				1	95	4				
F	91.18						93	9			
FF	89.80						1	88	9		
G	86.00							7	86	7	
GG	95.10								5	97	
H	89.00									11	89
Total	92.72	102	99	99	101	96	98	104	100	115	89

Note: Scorable events for each of questions 1.1 to 1.4 combined

B. Topics — no. variables = 5, WilksôLambda = 0.0099, F = 177.86, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	102									
B	92.93	5	92	2							
C	96.04		3	97	1						
D	98.99			1	98						
E	92.00					92	8				
F	89.22						91	11			
FF	96.84							95	3		
G	90.10							6	91	4	
GG	100.00									102	
H	84.00									16	84
Total	94.02	107	95	100	99	92	99	112	94	122	84

C. Performance expectations — no. variables = 5, WilksôLambda = 0.0072, F = 198.32, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	102									
B	91.92	6	91	2							
C	98.02		1	99	1						
D	92.93			5	92	2					
E	91.00				1	91	8				
F	89.22						91	11			
FF	96.94							95	3		
G	85.15							10	86	5	
GG	99.02									101	1
H	86.00									14	86
Total	93.03	108	92	106	94	93	99	116	89	120	87

D. Content — no. variables = 20, WilksôLambda = 0.0040, F = 42.7567, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	95.10	97	5								
B	89.90	7	89	3							
C	94.06		2	95	4						
D	98.99			1	98						
E	96.00				1	96	3				
F	88.24					1	90	11			
FF	89.90						2	88	8		
G	85.15							9	86	6	
GG	95.10								5	97	
H	88.00									12	88
Total	92.03	104	96	99	103	97	95	108	99	115	88

Table 6.49 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2005 SG Paper 1. Counts of correct classifications on the diagonal.

A. All scorable events — no. variables = 54, WilksôLambda = 0.0040, F = 14.55, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	81									
B	99.00	1	100								
C	94.95		2	94	3						
D	97.00			1	97	2					
E	89.11				4	90	7				
F	93.00						93	7			
FF	85.71						6	84	8		
G	83.17							9	84	8	
GG	92.16								7	94	1
H	84.85									15	84
Total	91.75	82	102	95	104	92	106	100	99	117	85

Note: Scorable events for each of questions 1.1 and 1.2 combined

B. Topics — no. variables = 6, WilksôLambda = 0.0123, F = 125.25, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	81									
B	98.01	2	99								
C	95.96		4	95							
D	95.00			5	95						
E	87.13				4	88	9				
F	87.00						87	13			
FF	97.96							96	2		
G	98.02							1	99	1	
GG	94.12								6	96	
H	83.84									16	83
Total	93.58	83	103	100	99	88	96	110	107	113	83

C. Performance expectations — no. variables = 5, WilksôLambda = 0.0115, F = 165.11, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	81									
B	98.02	2	99								
C	96.97		3	96							
D	95.00			5	95						
E	88.12				4	89	8				
F	87.00						87	13			
FF	98.98						1	97			
G	95.05							4	96	1	
GG	95.10								5	97	
H	83.84									16	83
Total	93.69	83	102	101	99	89	96	114	101	114	83

D. Content — no. variables = 15, WilksôLambda = 0.0092, F (135,7405) = 45.76, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	81									
B	95.05	3	96	2							
C	93.94		5	93	1						
D	94.00			4	94	2					
E	92.08				3	93	5				
F	90.00						90	10			
FF	95.92							94	4		
G	91.09							5	92	4	
GG	91.18								9	93	
H	84.85									15	84
Total	92.67	84	101	99	98	95	95	109	105	112	84

Table 6.50 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2005 SG Paper 2. Counts of correct classifications on the diagonal.

A. All scorable events — no. variables = 58, WilksôLambda = 0.0046, F = 11.18, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	6									
B	91.49		43	4							
C	95.05		4	96	1						
D	96.97			3	96						
E	96.00				3	96	1				
F	85.15					3	86	12			
FF	86.87						7	86	6		
G	85.00							8	85	7	
GG	94.95								5	94	
H	90.20									10	92
Total	91.33	6	47	103	100	99	94	106	96	111	92

Note: Scorable events for each of questions 1.1 and 1.2 combined

B. Topics — no. variables = 5, WilksôLambda = 0.0177, F (43,3760) = 122.42, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	83.33	5	1								
B	97.87		46	1							
C	98.02		2	99							
D	95.96			4	95						
E	90.00				7	90	3				
F	89.11						90	11			
FF	100.00							99			
G	94.00							6	94		
GG	90.91								9	90	
H	84.31									16	86
Total	92.97	5	49	104	102	90	93	116	103	106	86

C. Performance expectations — no. variables = 4, WilksôLambda = 0.0171, F(36, 32) = 171.7194, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	83.33	5	1								
B	97.87		46	1							
C	99.01		1	100							
D	96.97			3	96						
E	90.00				3	90	7				
F	90.10						91	10			
FF	98.99							98	1		
G	89.00							11	89		
GG	94.95								5	94	
H	85.29									15	87
Total	93.21	5	48	104	99	90	98	119	95	109	87

D. Content — no. variables = 15, WilksôLambda = 0.0127, F (135, 6472) = 36.13, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	83.33	5	1								
B	91.49	1	43	3							
C	96.03		4	97							
D	94.95			5	94						
E	89.00				5	89	6				
F	84.16					3	85	13			
FF	95.96						1	95	3		
G	90.00							7	90	3	
GG	95.96								4	95	
H	86.27									14	88
Total	91.45	6	48	105	99	92	92	115	97	112	88

Table 6.51 Classification matrix showing the number of scripts that were correctly and incorrectly classified – 2006 HG Paper 1. Counts of correct classifications on the diagonal.

A. All scorable events – no. variables = 50, WilksôLambda =0.032 , F = 16.45, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	97.35	147	4								
B	99.09		109	1							
C	96.81		2	91	1						
D	93.55			3	87	3					
E	92.62				2	138	9				
F	90.%					4	115	8			
FF	76.27						6	45	8		
G	87.06						1	8	74	2	
GG	92.59								3	50	1
H	84.91									8	45
Total	92.41	147	115	95	90	145	131	61	85	60	46

Note: Scorable events for each of questions 1.1 to 1.4 combined

B. Topics – no. variables = 6, WilksôLambda = 0.0125, F = 123.40, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	94.70	143	8								
B	100.00		110								
C	95.74		2	90	2						
D	92.47			3	86	4					
E	92.62				1	138	10				
F	98.43						125	2			
FF	74.58						7	44	8		
G	96.47							2	82	1	
GG	90.74								5	49	
H	92.45									4	49
Total	93.95	143	120	93	89	142	142	48	95	54	49

C. Performance expectations – no. variables = 5, WilksôLambda = 0.0116, F= 163.48, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	94.04	142	9								
B	99.09		109	1							
C	96.81		2	91	1						
D	91.40			5	85	3					
E	91.28					136	13				
F	94.49						120	7			
FF	67.80						12	40	7		
G	95.29							2	81	2	
GG	88.89								4	48	2
H	92.45									4	49
Total	92.41	142	120	97	86	139	145	49	92	54	51

D. Content – no. variables = 17, WilksôLambda = 0.0073, F = 37.42, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	96.02	145	6								
B	100.00		110								
C	95.74		2	90	2						
D	91.40			4	85	4					
E	90.60				3	135	11				
F	92.91					1	118	8			
FF	74.58						9	44	6		
G	91.76							4	78	3	
GG	72.22								9	39	6
H	79.25									11	42
Total	90.87	145	118	94	90	140	138	56	93	53	48

Table 6.52 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2006 HG Paper 2. Counts of correct classifications on the diagonal.

A. All scorable events — no. variables = 55, WilksôLambda = 0.0030, F = 14.64, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	46									
B	93.26	2	83	4							
C	98.70		1	76							
D	95.18			1	79	3					
E	90.22				2	83	7				
F	96.05						73	3			
FF	85.71						2	48	6		
G	81.30							3	87	17	
GG	92.80								7	116	2
H	90.36									19	178
Total	91.67	48	84	81	81	86	82	54	100	152	80

Note: Scorable events for each of questions 1.1 to 1.4 combined

B. Topics — no. variables = 5, WilksôLambda = 0.0131, F =156.22, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	46									
B	97.75		87	2							
C	100.00			83							
D	97.82				90	2					
E	86.73				3	85	10				
F	94.74						72	4			
FF	89.29							50	6		
G	92.59								100	8	
GG	100.00									126	
H	92.54									15	186
Total	94.87	46	87	85	93	87	82	54	106	149	186

C. Performance expectations — no. variables = 5, WilksôLambda = 0.0105, F = 169.30, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	46									
B	96.63	1	86	2							
C	92.77		2	77	4						
D	98.91				91	1					
E	86.73				3	85	10				
F	94.74						72	4			
FF	87.50						2	49	5		
G	88.89							2	96	10	
GG	97.62								2	123	1
H	89.05									22	179
Total	92.72	47	88	79	98	86	84	55	103	155	180

D. Content — no. variables = 17, WilksôLambda =0.0082, F (153,76) = 153.7615, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	10.00	46									
B	94.38	1	84	4							
C	90.36		3	75	5						
D	95.5			1	88	3					
E	86.73				6	85	7				
F	96.05						73	3			
FF	87.50						3	49	4		
G	87.96							3	95	10	
GG	96.83								3	122	1
H	89.55									21	180
Total	92.00	47	87	80	99	88	83	55	102	153	181

Table 6.53 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2006 SG Paper 1. Counts of correct classifications on the diagonal.

A. All scorable events — no. variables = 48, WilksôLambda =0.0050, F = 13.85, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	95.35	41	2								
B	95.24	3	120	3							
C	100.00			107							
D	97.98			2	97						
E	92.75				2	128	8				
F	92.62					2	113	7			
FF	84.13						5	53	5		
G	83.33							8	70	6	
GG	84.85								5	56	5
H	81.25									6	26
Total	92.16	44	122	112	99	130	126	68	80	68	31

Note: Scorable events for each of questions 1.1 and 1.2 combined

B. Topics — no. variables = 6, WilksôLambda = 0.0132841, F (54,4426) = 109.1743, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	95.35	41	2								
B	100.00		126								
C	100.00			107							
D	100.00				99						
E	92.75				6	128	4				
F	100.00						122				
FF	98.41						1	62			
G	94.05							1	79	4	
GG	89.55								7	60	
H	84.38									5	27
Total	96.59	41	128	107	105	128	127	63	86	69	27

C. Performance expectations — no. variables = 4, WilksôLambda = 0.1360, F (36,3254) =194.1858, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	93.02	40	3								
B	99.21		125	1							
C	100.00			107							
D	97.98			2	97						
E	92.03				6	127	5				
F	99.18						121	1			
FF	96.83							61	2		
G	95.24								80	4	
GG	83.58								11	56	
H	84.38									5	27
Total	95.46	40	128	110	103	127	126	62	93	65	27

D. Content — no. variables = 14, WilksôLambda = 0.0106, F (126,6568) = 42.4078, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	93.02	40	3								
B	96.83	2	122	2							
C	100.00			107							
D	98.99			1	98						
E	92.03				4	127	7				
F	93.44					3	114	5			
FF	84.13						4	53	6		
G	86.90							5	73	6	
GG	88.06								8	59	
H	84.38									5	27
Total	93.08	42	125	110	102	130	125	63	87	70	27

Table 6.54 Classification matrix showing the number of scripts that were correctly and incorrectly classified − 2006 SG Paper 2. Counts of correct classifications on the diagonal.

A. **All scorable events** — no. variables = 49, WilksôLambda = 0.0041, F = 14.24, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	6									
B	97.56		40	1							
C	93.26		4	83	2						
D	95.12			4	78						
E	95.92					94	4				
F	90.22					2	83	7			
FF	69.05						5	29	8		
G	88.89							4	80	6	
GG	93.85								7	122	1
H	95.71									9	201
Total	92.72	6	44	88	80	96	92	40	95	137	202

Note: Scorable events for each of questions 1.1 and 1.2 combined

B. **Topics** — no. variables = 5, WilksôLambda = 0.0159, F (45,3881) = 131.5546, p< 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	6									
B	100.00		41								
C	97,75		2	87							
D	98.78			1	81						
E	96.94				2	95	1				
F	98.91						91	1			
FF	50.00						7	21	14		
G	91.11								82	8	
GG	100.00									130	
H	94.79									11	200
Total	94.67	6	43	88	83	95	99	22	96	149	200

C. **Performance expectations** — no. variables = 4, WilksôLambda = 0.0145, F (36,3225) = 189.5877, p < 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	6									
B	97.56		40	1							
C	97.75		2	87							
D	95.12			4	78						
E	89.80				4	88	6				
F	100.00						92				
FF	57.14						7	24	11		
G	91.11							1	82	7	
GG	100.00									130	
H	94.79									11	200
Total	93.87	6	42	92	82	88	105	25	93	148	200

D. **Content** — no. variables = 15, WilksôLambda = 0.0101, F (135,6683) = 39.8935, p< 0.0001

Symbol	Percent correct	A	B	C	D	E	F	FF	G	GG	H
A	100.00	6									
B	92.68		41								
C	93.26		5	83	1						
D	93.90			3	77	2					
E	94.90				1	93	4				
F	94.57					2	87	3			
FF	71.43						8	30	4		
G	86.67							4	78	8	
GG	94.57								6	122	1
H	98.57									3	207
Total	93.74	6	46	86	79	97	99	37	88	133	208

rather than by the use of critical process. Descriptions of the performances standards (Section 6.2.2) in this thesis show that students who receive the same symbol in what is considered, by virtue of the same qualification, to be the same examination (e.g., HG Biology 2005 and HG Biology 2006), demonstrated different competencies at the same level of mastery. Performance standards had different meanings between years because the content standards in each examination were different and because performance standards take their meaning from the content standards. Classifications based on criteria which define performance standards other than by aggregate mark might yield different results. They would depend on where the cut-scores are set within and between years and on the content standards of a particular examination. This analysis has implications for the assumption made about the equivalence of examinations, and will be discussed in Section 6.4.

The 2006 performance data, where marks for Paper 1, Paper 2 and the aggregate mark were available from the same sample of students, are shown in Figures 6.26 and 6.27. The pattern for both HG and SG is that students tended to do better in Paper 1 than they did in Paper 2. Should the cut-scores have been the same for the papers within each set? The marks for Paper 1 and Paper 2 were summed to obtain the aggregate mark, thus enabling the question of whether it makes a difference that one question paper was more difficult than the other question paper to be answered. When Umalusi adjusted marks before certification, assuming that the marks were adjusted,[222] did the adjustment occur per paper or per aggregate?

Section 6.2 has shown variability in student performance, depending on the content standards and the profile of a question paper. With respect to the structural components of a question paper, more able students generally performed according to the design of each question papers; weaker students performed relatively better when they could choose the correct answer, rather than construct a free-response answer, and were less able than stronger students to construct longer answers that shorter answers to questions. How generalizable these results might be to the entire population is unclear due to the process by which the scripts were selected for this study (Chapter 5, Section 5.1.3). Despite that the author followed rigorous procedures to validate the PET later used to classify performance expectations (Chapter 4, Section 4.3.2), others might have classified examination questions differently with respect to cognitive demand. Therefore, the examiners' classifications for two of the eight question papers according to cognitive demand were cross-referenced with the PET in Chapter 4 (Section 4.3.3.2) and they will again be cross referenced in the discussion of question difficulty according to performance below.

[222] The first SC mark adjustments to be made public were those of the 2010 NSC examinations.

		Paper 2									
		A	**B**	**C**	**D**	**E**	**F**	**FF**	**G**	**GG**	**H**
Paper 1	**A**	A[a] 45[b]	A 48 B 26	B 23	B 4 C 5						
	B	B 1	B 14	B 24 C 20	C 35	C 88 D 4	D4				
	C		C 1	C 15	C 15 D 20	D 24	D 3 E 4	E 1	E 1		
	D			D 1	D 10	D 20 E 21	E 28	E 7	E 2 F 3	F 1	
	E				D 2 E 1	E 11	E 21 F 11	E 1 F 33	F 29 FF 6	F 7 FF 22	FF 2 G 3
	F						F 5	F 2 FF 5	F 5 FF 37 G 4	F1 FF 17 G 24	G 23 GG 3 H 1
	FF							FF 5	FF4 G 8	GG 22	G 2 GG 18
	G							FF 1 G 1	G 8	G 4 GG 17	GG 42 H 12
	GG								G 1	GG 10	GG 10 H 33
	H									H 1	H 52

Note: a Aggregate mark (Paper 1 plus Paper 2)
b Number of candidates

Figure 6.26 Relationship between symbols achieved in HG Paper 1, HG Paper 2 and the aggregate mark used for certification, 2006 sample scripts only. (n = 975)

		Paper 2									
		A	B	C	D	E	F	FF	G	GG	H
Paper 1	A	A[a] 4[b]	A 4 B 12	B 19	B 3 C 1						
	B	A 1 B 1	B 24	B 36 C 19	C 29	C 10 D 3	D 3				
	C		B 1	C 11	C 26 D 14	D 45	D 6 E 2	E 1	E 1		
	D			C 2 D 2	D 9	D 16 E 17	E 30	E 11	E 5 F 2	F 4	GG 1
	E					E 7	E 22 F 18	F 17	F 28 FF 1	F 11 FF 22	FF 7 G 6
	F						F 8	F 6 FF 5	F 5 FF 19 G 1	FF 25 G 22	G 19 GG 3
	FF						FF 3	FF2	FF 2 G 8	G 18	G 4 GG 26
	G							G 1	G 8	G 7 GG 11	GG 45 H 12
	GG								G1	GG 6	GG 5 H 55
	H									GG 1 H 3	H 28

Note: a Aggregate mark (Paper 1 plus Paper 2).
b Number of candidates.

Figure 6.27 Relationship between symbols achieved in SG Paper 1, SG Paper 2 and the aggregate mark used for certification, 2006 sample scripts only. (n = 881)

While this section provides details of the relationships that exist between student performances and content standards, future detailed psychometric analyses might yield more insight into the precise nature of these relationships. This objective is, however, beyond the scope of this thesis. In their surveys of the enacted curriculum Porter and Smithson (2001) tried to include a third dimension, called mode of presentation (of instruction),[223] to the content matrix but found that it was difficult to use and that did not correlate well with student performance. These authors suggested that a third dimension, mode of representation (of instruction),[224] might contribute a better understanding to student performance. This author suggests that a future study might examine aspects of the scorable events in three dimensions: topic, performance expectations and the various measures of structure used in this thesis.

6.3 Difficulty of examinations

Given that worldwide the diversity of students within schools is increasing in many contexts, there is a need for more research into what makes science assessments difficulty in various contexts (Penfield & Lee, 2010). Comparing the difficulty of examinations is multifaceted (Schwille, 1996), and requires an understanding of the educational and examination systems in which each examination is located (Britton et al., 1996b). Britton et al. (1996a, 1996b) listed, with reasons, a number of different examination characteristics that ought to be considered when judgments are made about the difficulty of examinations. They noted that some combinations of examination characteristics could lead to contradictory judgments (Britton et al., 1996b). For example, contextual issues around examinations such as, for example, teaching and prior coursework, whether students had learned the work and were simply recalling it in an examination or not, and how much coaching students had for an examination, are extremely important (Eubanks & Eubanks, 1996; Schwille, 1996). To avoid inaccurate generalizations about the difficulty of examinations it is important that the criteria used to comment about difficulty of examinations, be explicitly justified, especially when used for comparative purposes (Britton et al., 1996b).

In Chapter 2 (Section 2.1), the educational and examination systems in which the SC Biology examinations were embedded were discussed and efforts used to determine the difficulty of SC examination question papers were described in Chapter 5 (Section 5.2.3.4). In this section, the

223 "The distinctions included: exposition — verbal and written, pictorial models, concrete models (for example, manipulatives), equations or formulas (for example, symbolic), graphical, laboratory work, and fieldwork" (Porter & Smithson, 2001, p. 9).

224 "[F]or example, written, symbolic, or graphic representation" (Porter & Smithson, p. 10).

author approaches the difficulty of the SC Biology examinations from two viewpoints, namely, the question papers and student performance in eight of these question papers.

6.3.1 Question papers

International studies of mathematics and science school-leaving examinations argued how complicated it is to judge the comparative difficulty of examinations using only question papers (Gandal, 1994; Britton et al., 1996a). Gandal (1994) argued for knowing information about how the examinations were graded, how students were prepared for the examinations, the scope of the examination system, and how much each examination counted towards students future expectations when comparing examinations. Britton et al. (1996a, 1996b) did not use student data to base the comparisons of examinations which they conducted on the mathematics and science school leaving examination of seven countries. These authors ranked a given country's examinations according to examination characteristics which they argued were important in determining the difficulty examinations, and presented their results for all countries that they studied in rank order of their perceived difficulty. It was not possible to order these countries in terms of overall difficulty because consideration of each of the examination characteristics led to different, sometimes contradictory, conclusions (Britton, Hawkins, & Gandal, 1996). Comparable data about the SC Biology examinations is combined with the data from Britton et al. (1996a) (Table 6.55). Table 6.55 will not be discussed in detail here because discussions of the effects of many of the examination criteria listed in this table have been woven into previous chapters of this thesis, as and when appropriate. Its purpose here is simply to supply a contrast for SC Biology examinations. Generally, the secondary school examinations studied in Britton and Raizen (1996) all determined the future career prospects of students. Black (1996) questioned how in one subject, like Biology, two examinations, one more than eight hours long and the other just more than two hours long, could fulfil this same aim. A discussion of the optimal length of the SC Biology examinations will be discussed in Section 6.5.

The difficulty profile of South African SC Biology examinations was approximately midway between the range of practices observed by Britton et al. (1996a) (Table 6.55). One exception was that post-2000 the national DoE examination was split across two papers, each two hours long and covered approximately half of the content. Given that all of the question papers in this study included questions labelled as Memorize questions, and that there was less work to memorize for each of the two papers (because topics were split across two question papers), it could be argued that the national DoE examinations were easier from 2001 onwards. Similarly, it might be argued that the IEB which continued with one question paper with all the content potentially examined in that paper consequently had more difficult examinations. Proponents

Table 6.55 Features of secondary school exit level Mathematics and Science examinations (1991, 1992, 1993) of eight countries (Britton et al., 1996a, p.50), including South Africa. Data for the South African SC HG and SG Biology examinations (1994 to 2007), for DoE and IEB is in bold italics.

Feature	Examining body	Practice
How long is the exam?	England & Wales	6 to 9 hours, mostly 8 hours
	Israel	5 to 5.5 hours
	France, Germany, Sweden	3 to 4.5 hours, mostly over 3.5 hours
	DoE – 1994 to 2000, IEB – 1994 to 2007	***3 hours***
	DoE – 2001 to 2007	***2 x 2 hours***
	Japan, United States	2.5, 3 hours
Do students choose among questions?	Germany, Japan, Sweden	No choices
	France, United States	Considerable, but only in some subjects
	DoE – 1994 to 2000 (HG only), IEB – 1994 to 2007	***Choice of questions***
	DoE – 2001 to 2007	***No choices***
	England & Wales, Israel	Extensive, all subjects
What item types are used?	England & Wales, Israel	Laboratory practicals
	France, Germany	Free-response only
	England & Wales, Israel, Japan, Sweden	Free-response with some multiple choice
	DoE, IEB – 1994 to 2007	***Free-response with some multiple choice***
	United States	Least free-response, most multiple choice
How broad is the examinationsôtopic coverage?	England & Wales, Germany, Israel, United States	Broad, many topics
	DoE – 1994 to 2000, IEB – 1994 to 2007	***Variable, broad to fewer topics***
	DoE – 2001 to 2007	***Broad, the same across all years***
	France, Japan, Sweden	Significantly fewer topics
What student performances are expected?	France, Sweden	Recall only, 20%
	Israel, Germany	Recall only, 33.3%
	England & Wales, Japan	Recall only, 40%
	United States	Recall only, 50%
	DoE, IEB – 1994 to 2000	***Knowledge, 60% (HG)*** ***Knowledge, 75% (SG)***
	DoE – 2001 to 2007	***LOCS, 60% (HG)*** ***LOCS, 75 to 80% (SG)***
	IEB – 2001 to 2007	***LOCS, 60% (HG)*** ***LOCS, 75% (SG)***

Note:
1. For each facet of difficulty, countries are listed in decreasing order of difficulty according to Britton et al., (1996).
2. LOCS = knowledge, comprehension (See Chapter 4, Section 4.3.1.3).
3. LOCS = knowledge, comprehension, application (See Chapter 4, section 4.3.1.3).

who argued in this way might have attributed the increased national DoE HG and SG pass rates post-2000 (Chapter 2, Table 2.10) to easier question papers, but the pass rate in IEB examinations also increased post-2000 in one question paper.[225] Another enquiry might note that South African HG and SG policies allowed for up to 80% recall only questions, which was a much higher proportion than was found in France, Sweden, Israel, Germany, England, Wales and the USA. An implication of this data could be that South African SC Biology examinations were easier than corresponding examinations of the other countries. Such a conclusion should be tempered with caution given that the South African CBS was unclear as to what exactly constituted a recall question. The ambiguities created by the lack of explicit clarity of cognitive demand categories in South African policies were discussed in Chapter 4 and previously in this chapter.

6.3.2 Student performance (Appendices 6.59 & 6.60)

Understanding the sources of difficulty of test items makes it easier to set tests but it is difficult for even experienced practitioners to predict item difficulty for populations which they are familiar (Scheuneman, Gerritz & Embretson, 1991). Item analysis can provide information about the difficulty, the discrimination, and the differential item functioning of the items comprising a test. If conducted after a test is written, analyses of student performance can give an indication of the full range of difficulty of the items in a test (Livingston, 2006). While there are a number of complicated statistical methods used in item analysis (e.g., Downing, & Haladyna, 2006), the simplest is the average mark achieved by a group of students on a test item; this is known as the item difficulty level (Nitko & Brookhart, 2007). Item difficulty level is a function of the complexity of the task required by the item (Embretson, 1983; Sax, Eilenberg & Klockars, 1972), the nature of the instruction which precedes the task, and the ability of the students responding to the task (Nitko & Brookhart, 2007). Therefore, using only the average mark achieved for a test can be misleading, especially if the group of test takers was unusually strong or weak (Livingston, 2006).

The reader is reminded that in this study the number of student scripts selected for analysis was not in proportion to the number of students who achieved each performance standard or symbol (Chapter 5, Tables 5.1 & 5.2). The results that follow thus represent samples of similar numbers of students within each symbol. This mode of selection means that these samples of students are not necessarily representative of the full examinee populations which were skewed towards the lower symbols for Biology. Difficulty, as experienced by students of differing

225 It may have been that the increases in pass rate were the result of statistical moderation. There was no public data by which the author could check what statistical adjustments were made.

levels of mastery, has already been discussed (Section 6.2.2.2). Comparative levels of question difficulty are used here to extract first insights into the features that might make particular SC Biology questions more or less difficult than other questions. Comparison between categories of content was not made because many content categories were represented by few or no scorable events.

For each of the eight question papers for which student performance was analyzed here, summary descriptions for the questions in which various samples of students scored more than 80% (i.e., easiest questions) and less than 20% (i.e., most difficult questions) are given in Appendices 6.59 and 6.60. Several of the HG questions in which students scored less than 20% related to practical work, highlighting that very little practical work appears to have been taught or learned during this time and the potential problems of testing practical Biology work in a paper-and-pencil manner, as discussed previously (Section 6.1.1.1a). Many of the questions for which each sample of students received a mean mark of less than 20% , carried more than 2 marks, which meant they required longer answers. Further inspection of the question descriptions given in Appendices 6.59 and 6.60 provided no additional consistent, logical features which might explain the differences in student performance that accompanied these questions, although there are indications that HOCS questions might be more difficult. The absence of salient features points to the complex nature of SC Biology questions (and their answers) that has emerged several times during this study. It highlights the need for further investigation in future studies of the relationships between variables, and of the synergy between the variables, that comprise SC examination question papers, especially because cognitive complexity has been shown to affect student perceptions of task difficulty (Robinson, 2001).

Student performance with respect to the question characteristics used in this study showed a range of achievement in different scorable events (Figures 6.28 & 6.29), yet no consistent patterns in difficulty could be observed. Generally, scorable events which required students to choose the correct answer rather than to construct a free-response answer, were easier and showed a smaller range in difficulty, and writing extended answers to questions was more difficult (Figures 6.30 & 6.31). Almost all questions which required students to choose the correct answer were classified as LOCS questions in this study. No consistent patterns in the difficulty of scorable events were discernable, with respect to topics (Figures 6.32 & 6.33).

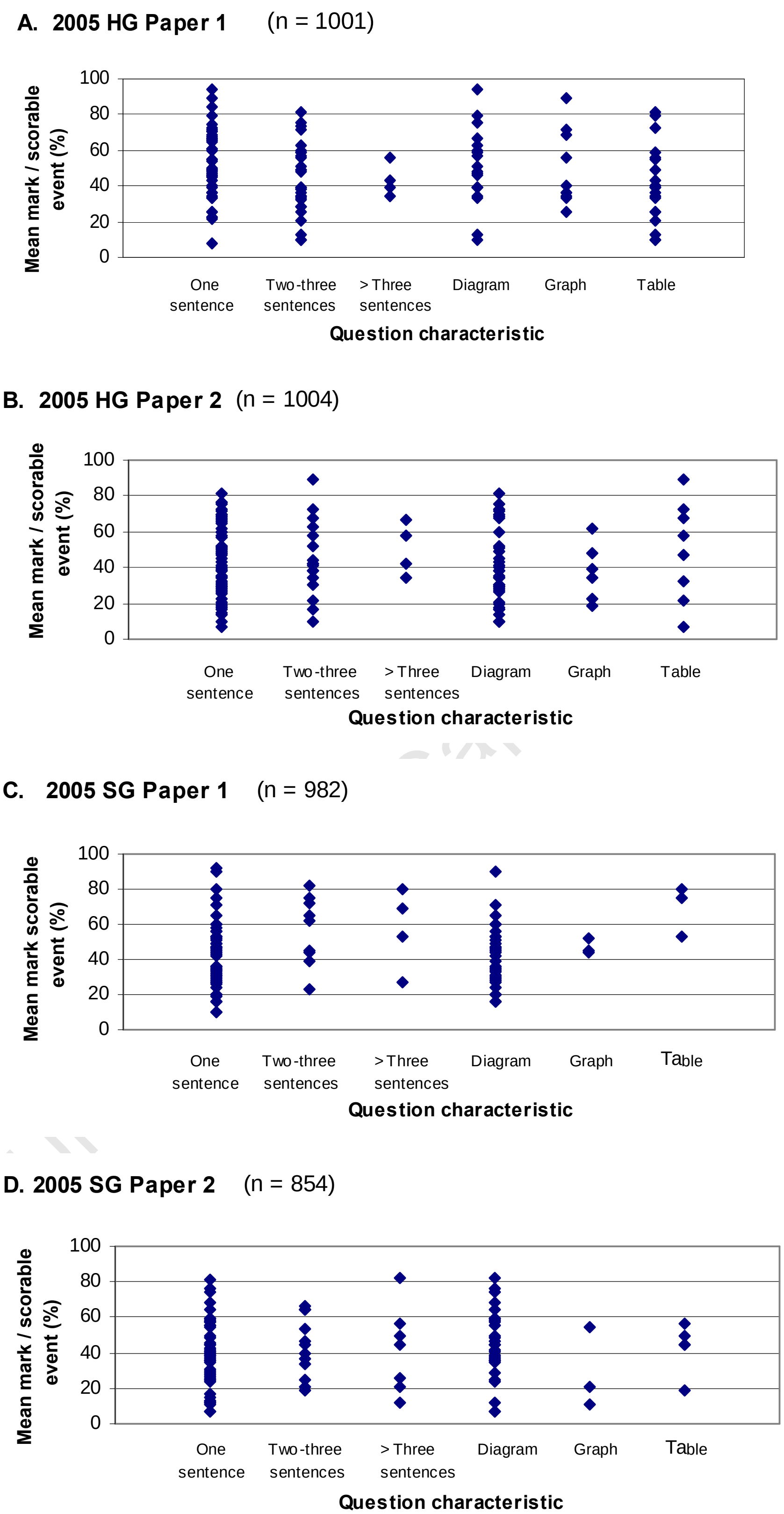


Figure 6.28 Mean mark per scorable event for 2005 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by question type.

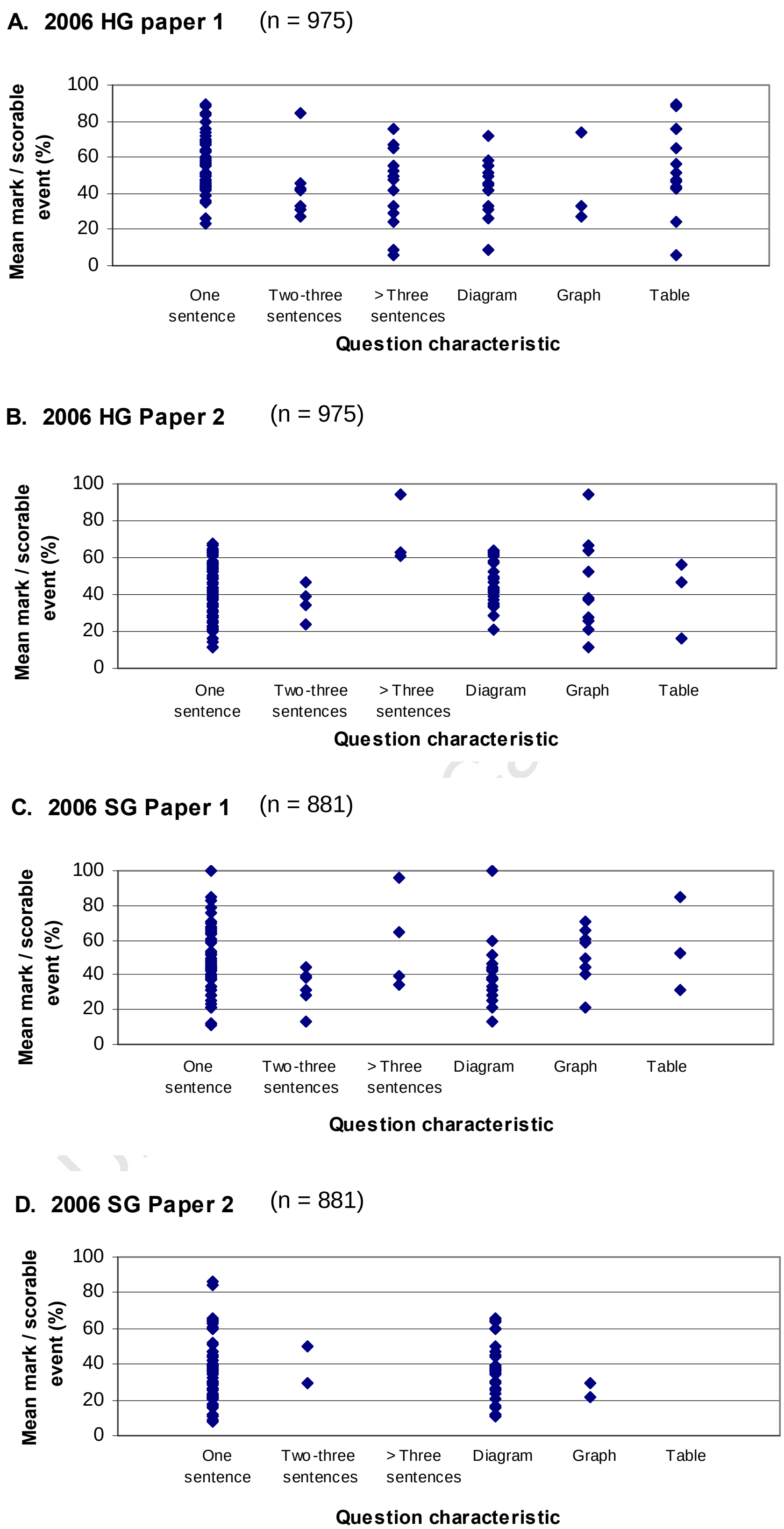
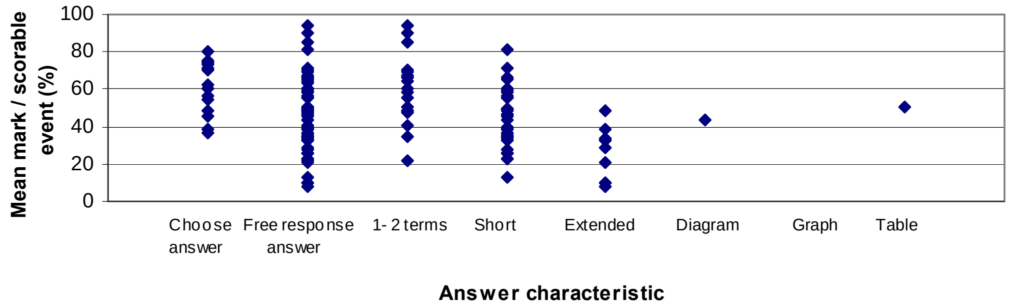
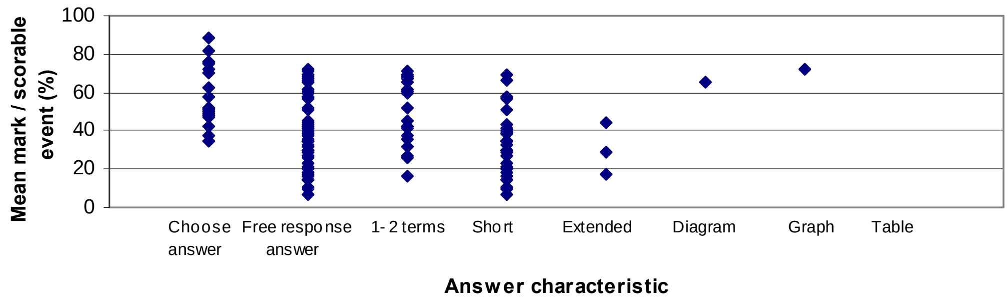


Figure 6.29 Mean mark per scorable event for 2006 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by question type.

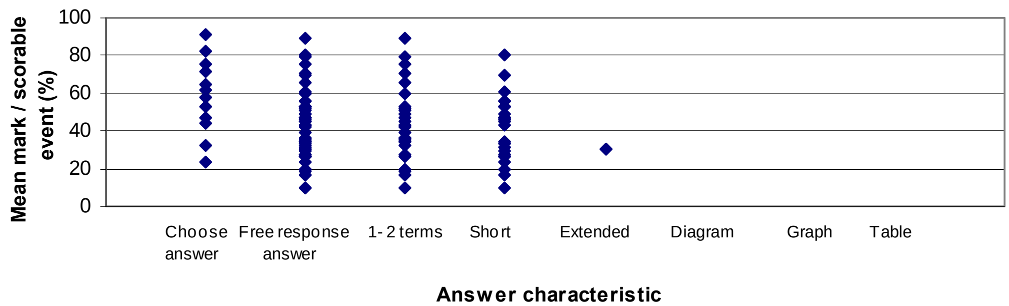
A. 2005 HG Paper 1 (n = 1001)



B. 2005 HG Paper 2 (n = 1004)



C. 2005 SG Paper 1 (n = 982)



D. 2005 SG Paper 2 (n = 854)

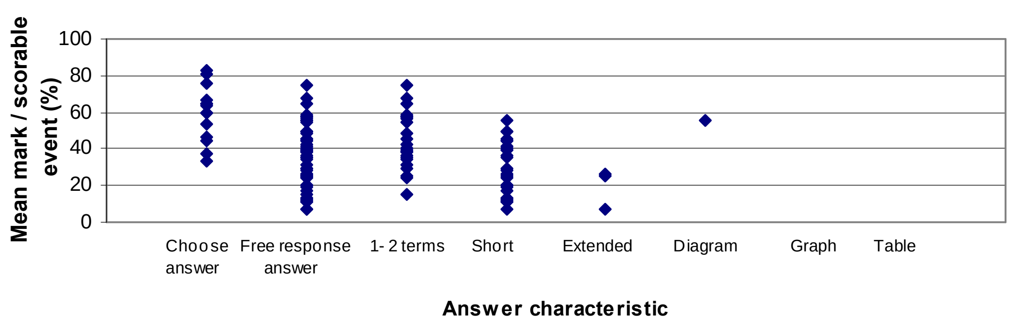


Figure 6.30 Mean mark per scorable event for 2006 HG Paper 1 and Paper 2 and SG Paper 1and Paper 2, by answer type.

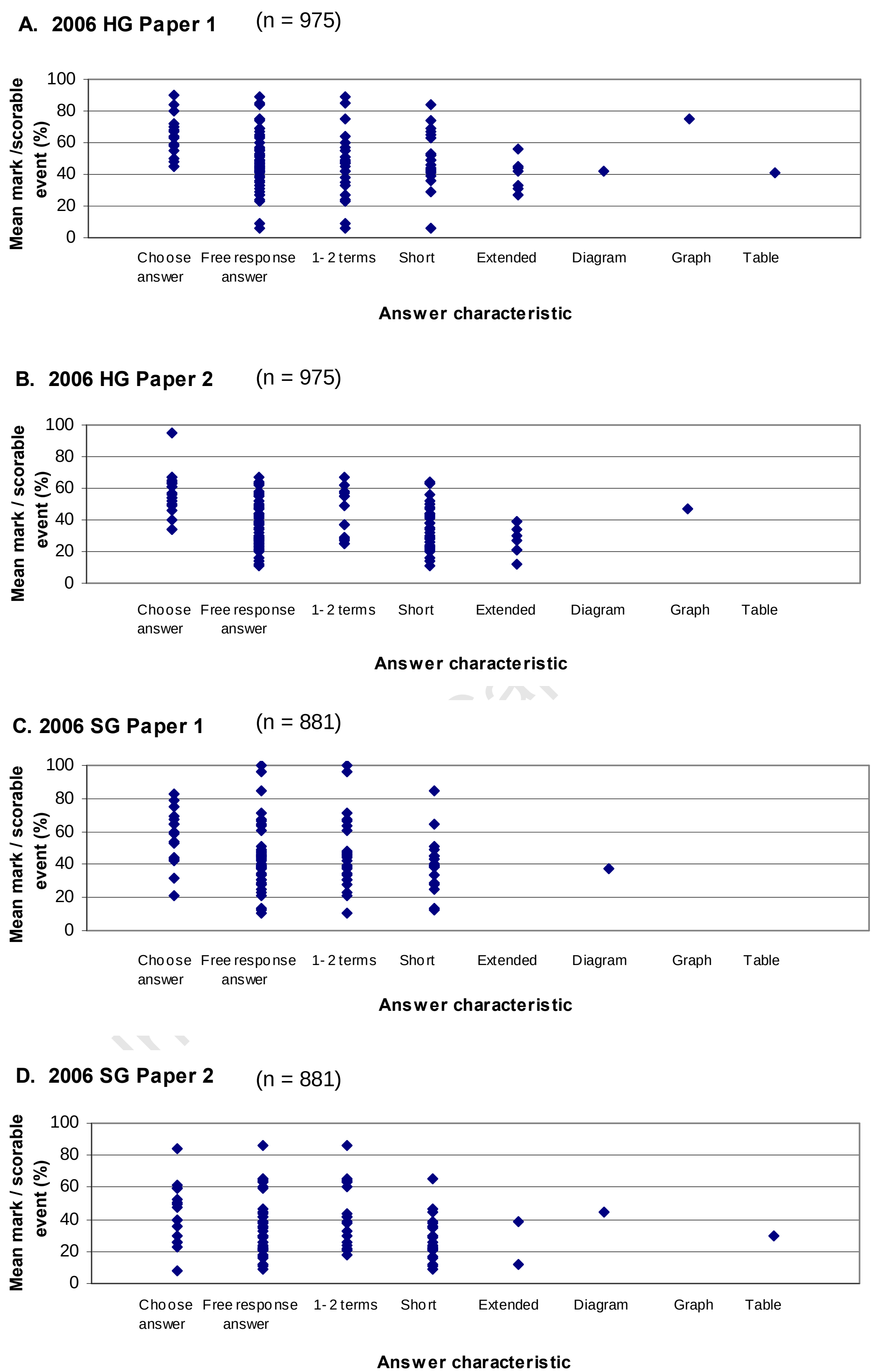
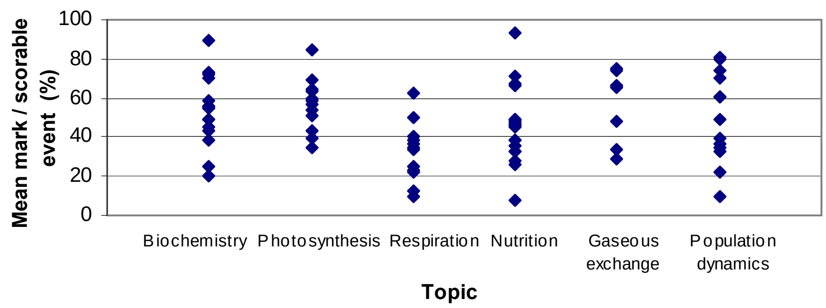
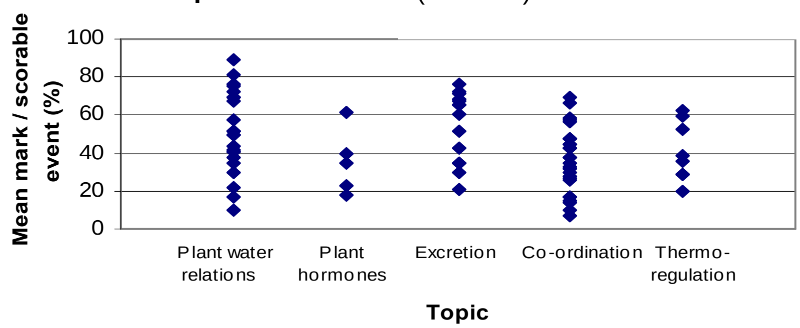


Figure 6.31 Mean mark per scorable event for 2006 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by answer type.

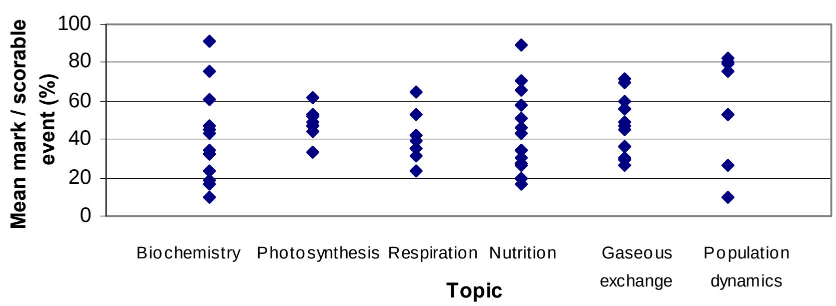
A. 2005 HG Paper 1



B. 2005 HG Paper 2 (n = 1004)



C. 2005 SG Paper 1 (n = 982)



D. 2005 SC Paper 2 (n = 854)

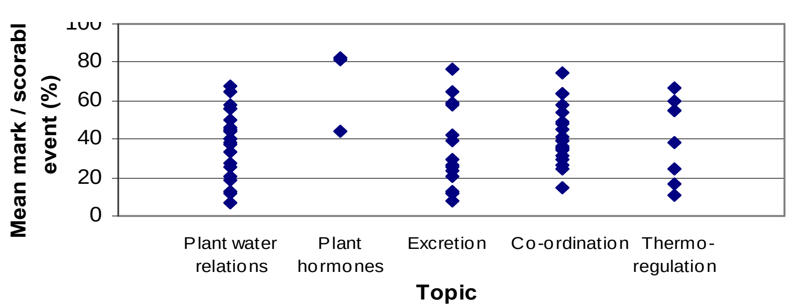


Figure 6.32 Mean mark per scorable event for 2005 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2 by topic.

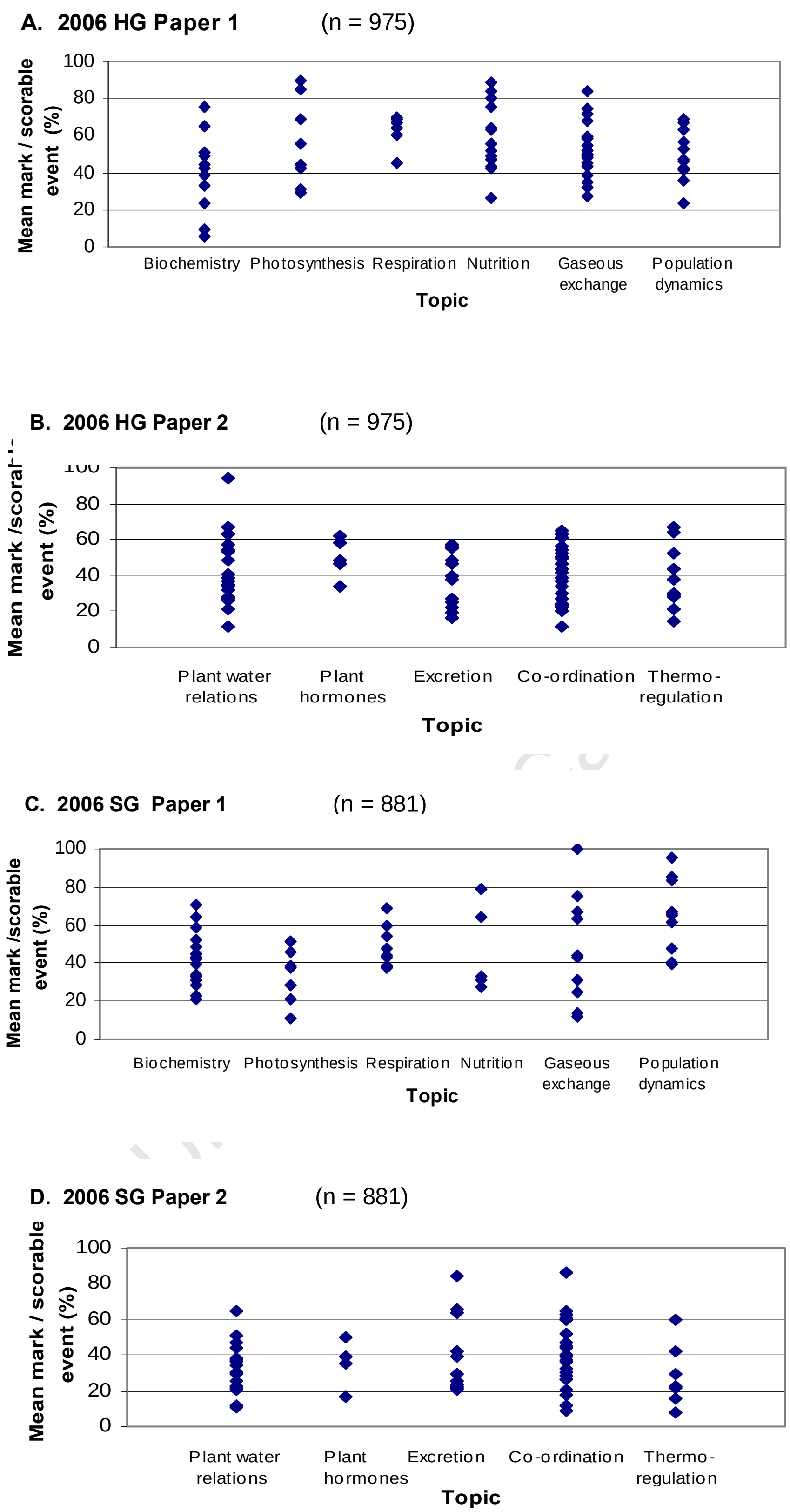


Figure 6.33 Mean mark per scorable event for 2006 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by topic.

On its own, cognitive demand has little value in predicting the difficulty of a test item (Scheuneman et al., 1991). Scorable events that required students to recall knowledge (PET category Memorize) showed an enormous range of difficulty (Figures 6.34 & 6.35). This variation suggests that memorizing knowledge is not necessarily easier than applying knowledge. The data from this study shows that several levels of difficulty exist within each category of cognitive demand, which constitutes a new level of complexity. In addition, scorable events that required students to apply their knowledge to contexts outside of what they learned, tended to be more difficult than scorable events that required students to answer within the learned contexts (Figures 6.34 & 6.35). The results of this study supports the views that difficulty operates within each of the cognitive levels (Sousa, 2006), and that difficulty also may increase with increased complexity (Bloom et al., 1956). The relationship between complexity and difficulty and their combined effect on student performance deserves to be explored in future research, especially since it has been suggested that complexity explains item difficulty (Scheuneman et al., 1991).

Similar patterns in cognitive demand were not obvious within the analysis of the difficulty of questions classified by cognitive demand as recorded by the examiners of two of the question papers (Figure 6.36). This extreme contrast reflects the effect of measuring cognitive demand with different tools in comparative studies (Chapter 4, Section 4.3.2).

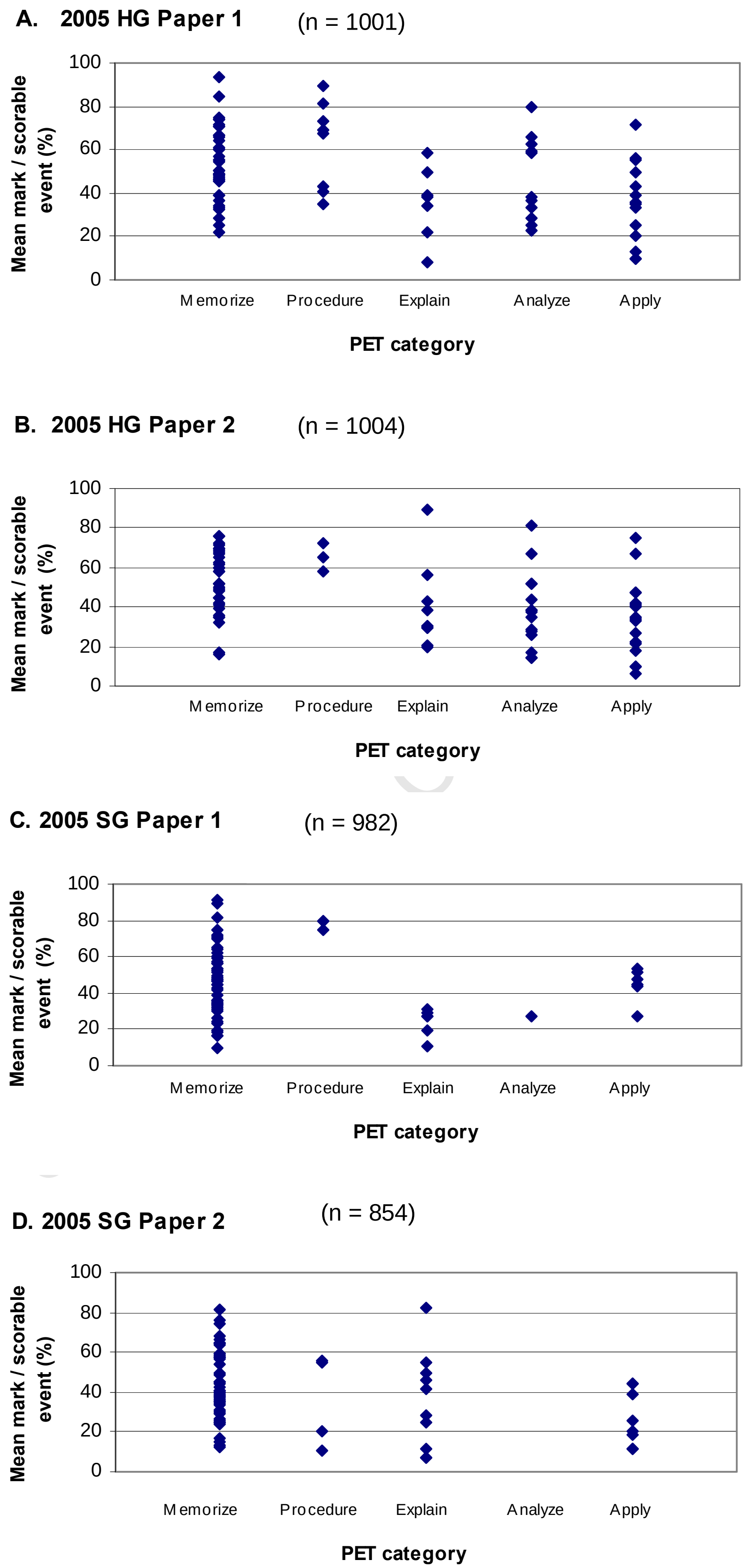


Figure 6.34 Mean mark per scorable event for 2005 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by PET category.

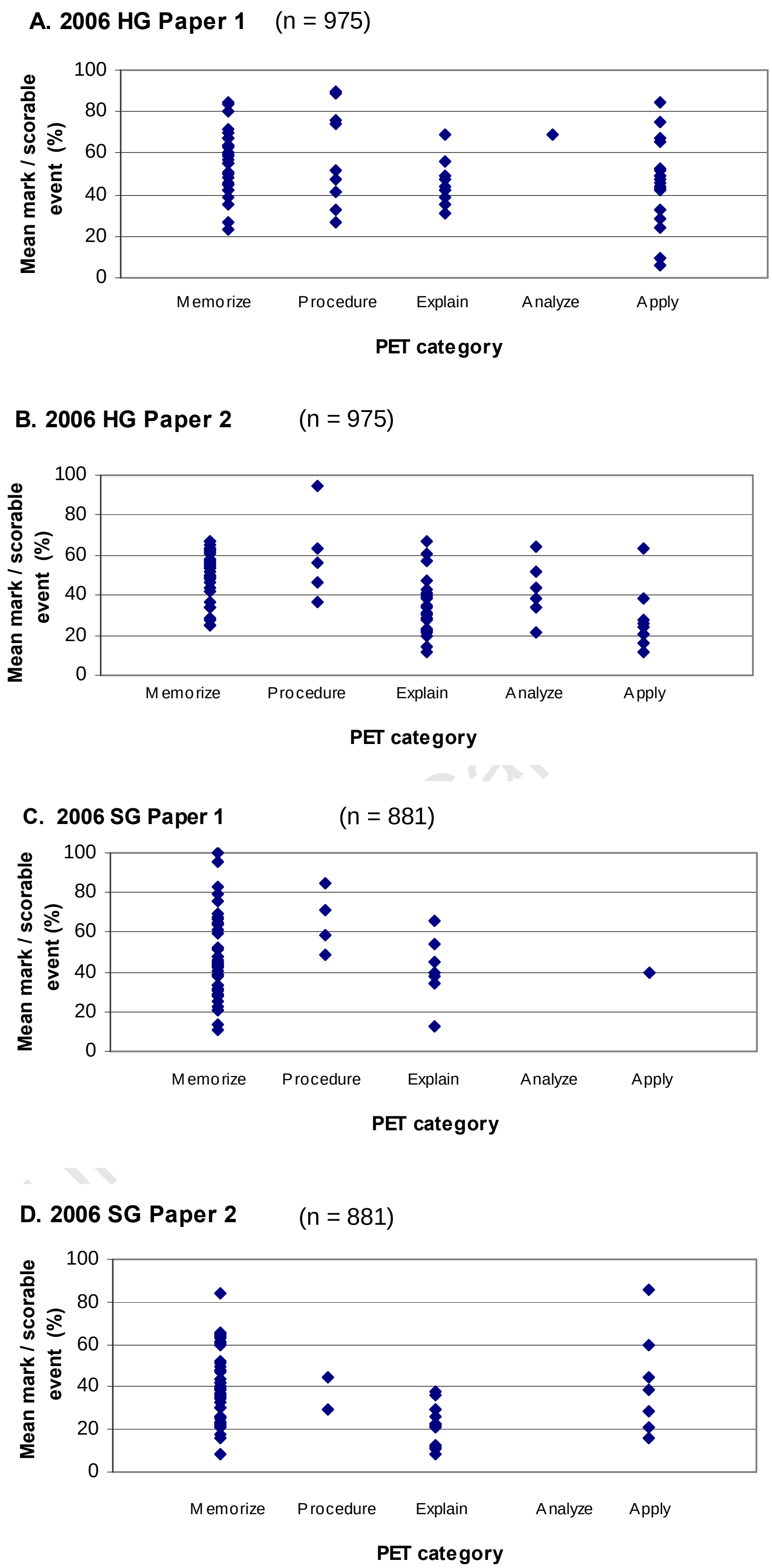
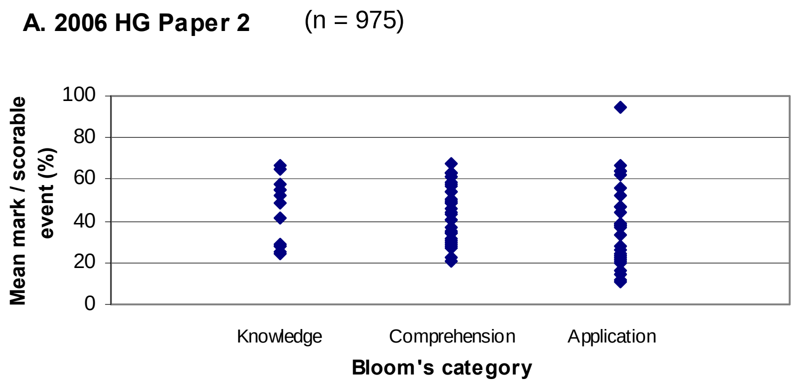


Figure 6.35 Mean mark per scorable event for DoE 2006 HG Paper 1 and Paper 2 and SG Paper 1 and Paper 2, by PET category.



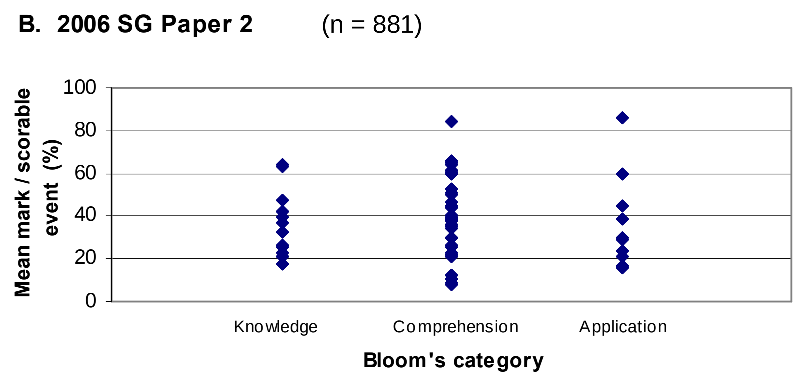


Figure 6.36 Mean mark per scorable event for DoE 2006 HG Paper 2 and DoE 2006 SG Paper 2 by BTEO categories.

6.4 Equating of examinations

Equating of tests is an attempt to ensure that claims of comparable evaluative judgments made about students, is fair. When evaluative judgments are to be made about the equitability of students' examination scores from different versions of the examination, it is important for validity reasons that that equivalence (or lack of equivalence) be established between different examinations which purport to be consistent in their classifications of students (Chapter 3). Put more simply, equivalence means that past and present students should be required to know and perform tasks of similar complexity to achieve the same mark (Nitko & Brookhart, 2007). Brennan (2006b) described three ways, *equating, linking* and *scaling* (Chapter 3, Section 3.4) by which equivalence, or its violation, can be established. The extent to which these methods are to South African SC Biology examinations equivalence practices will now be examined.

The question papers analyzed in this study comprised various different profiles with different content standards (Section 6.1.2), and *equating* as described by Brennan (2006b) would therefore not apply

Each year students were certified as having a level of mastery, a performance standard, represented by a symbol from A to H, on either the HG or the SG. Each of the symbols A to H represented an aggregate mark for a student (Chapter 2, Section 2.1.1). The cut-scores which separated each of the symbols remained constant between years, but differed between HG and SG (Chapter 2, Section 2.1.1). In the absence of consistent content standards, new/different cut-scores must be set for each test to accommodate differences in content standards. Therefore, *linking* as described by Brennan (2006b) did not apply.

Statistical adjustments made to the SC examination raw scores have been purported to ensure that the allocation of a student to a particular category symbol was fair and consistent within and between years (Chapter 2, Section 2.1.3.3). That is, students who received the same symbol within a year (different examining bodies, different question papers), and between different years (different examining bodies, different question papers; same examining bodies, different question papers) were presumed equivalent. This type of adjustment is similar to some forms of *scaling* but scaling is meant to confer meaning of test scores which teachers and other test score users may invoke to make inferences about students' learning or abilities.

Statistical adjustments in 2010 were such that "[i]n some instances different levels of adjustments are effected in different parts of the mark distribution" (Umalusi 2011, no page number). A risk associated with a focus of the statistical moderation on ensuring equivalence between years (Fatti, 2006) is that adjustments may not have been fair to all students within a year. For example, the results of this study has shown that differences in the performance of students within each of the 2005 and 2006 SC Biology examinations were simply different degrees of emphasis on particular content categories in each examination. If the marks of all students within one of these years were increased or decreased, the adjusted mark would have either inflated or deflated all the marks within the year. However, if the marks of only the students who achieved lower marks (e.g., the majority of SG students) were increased, these students would have been advantaged relative to students who achieved higher symbols in the same year, for the same suite of competencies. Can standardizations between years or even between examining authorities without integration of explicit performance standards associated with each mark distribution claim fairness? Unfortunately, no score adjustment data could be obtained for 2005 and 2006 SC Biology examinations.

How does the above discussion of equivalence impact the assumptions that were made by the certifiers, and the users of the SC Biology results? Certifiers assumed that HG and SG examinations, while sufficiently different to award a HG or a SG qualification, were sufficiently similar to allocate students who failed on the HG, a notional SG pass (Chapter 2, Table 2.2; Figure 6.37). A comparison of the descriptions of the performance standards in the 2005 and 2006 HG examinations showed that HG students who received converted SG pass symbol exhibited different suites of abilities to SG students who wrote the comparable SG examinations and received the same SG symbol (Section 6.2.2). Similarly, some universities assumed an equivalence between HG and SG symbols for selection purposes (Chapter 2, Table 2.3) that was not confirmed in this study (Figure 6.37). Equivalence of the SC examinations assumed equivalence of particular performance standards, but performance standards are not comparable if the corresponding content standards of the examinations being compared are not identical, as performance standards have meaning in relation to specific content standards. The standardization of SC marks claimed to "deliver a relatively constant product to the market: universities, colleges and employers (Fatti, 2006, p. 46). This study showed that different students who received the same level of SC qualification in 2005 and in 2006 has demonstrated different competencies, because the content standards of each examination was quite different.

Assumptions of equivalence between some HG and SG grades are further challenged by the fact that HG students who achieved a failed HG symbol and a converted SG symbol, had to work harder than students who achieved the same SG symbol by writing the SG question paper. For example, a HG student who failed on HG with 140 marks out of 400 was awarded a SG 'F' pass. In contrast, a SG pupil needed between 100 and 119 marks to be awarded an SG 'F' pass. As both the HG and SG examinations were written in the same period of time, HG students potentially had to work harder per unit time than SG students to be awarded the same Grade and symbol pass. Indeed this same inadequacy for comparability, differing effort per unit time, had similar implications whenever within years the question papers from different examining bodies of the same Grade (HG or SG) carried different number of total marks (Section 6.1.1). In this argument the author herself assumed that a mark carries consistently the same value between examinations of the same examining bodies, the same Grade and between years. This assumption was not vindicated (Section 6.1.1).

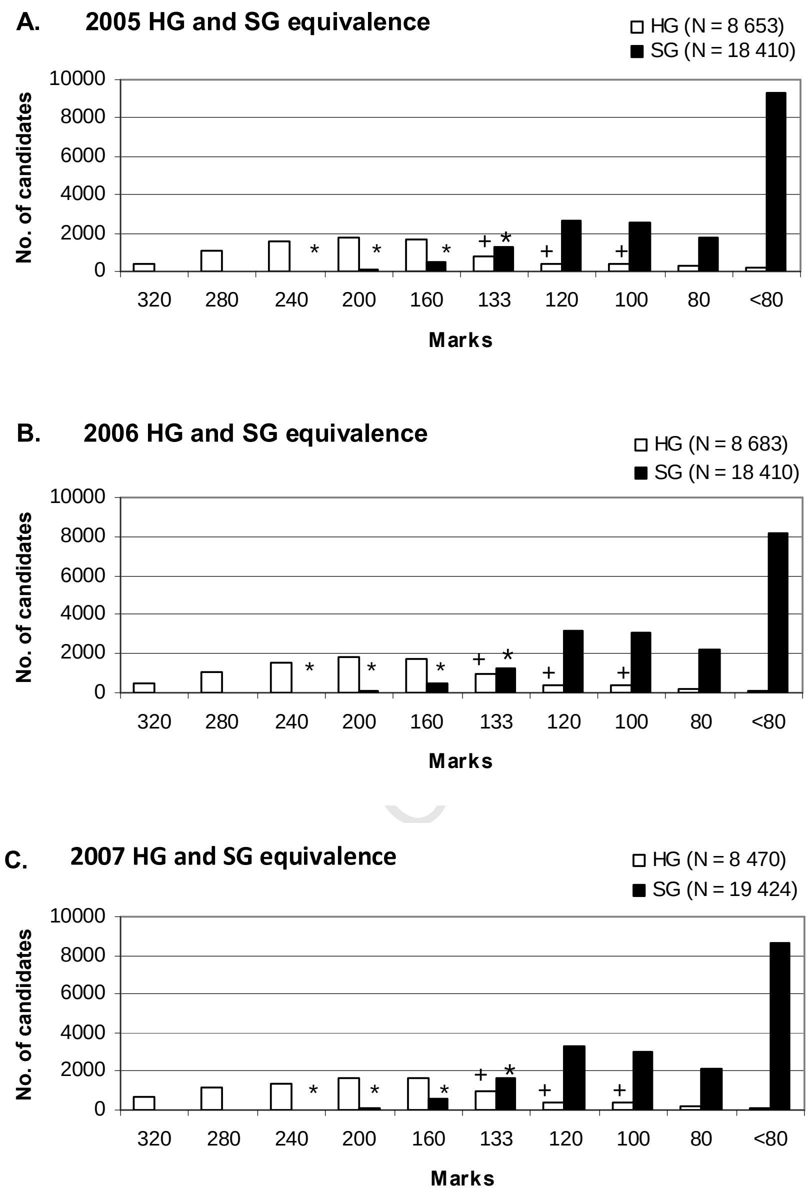


Figure 6.37 WCED HG and SG student performance according to aggregate marks, 2005 to 2007. (+ indicates HG failures converted to SG passes; * indicates SG passes that carried the same point ratings as the HG equivalent for admission to some universities)

What should a consistent value for a mark be defined as in SC Biology examinations? Should the value be defined in terms of “one mark per one fact”? During the period 2001 to 2007 this equating of one fact and one mark was the predominant practice in the national DoE HG and SG examinations. Or, should the value of a mark be defined by the complexity and mental effort required by a question? This option was practiced, in part, by the IEB 2001 to 2007 HG examinations. Clearly, a conversation about the value of a mark needs to take place to start to address equivalence between comparable examinations set by different examining bodies, and between comparable examinations set in different years.[226,227]

Performance standards derive meaning from the content standards to which they refer (Chapter3, Section 3.5.3). Because each distinct question paper examined distinct suites of content standards, descriptions of the same performance standard differed between years (Section 6.2.2). This variability is further exacerbated when the policies of participating examining bodies are different (e.g., the national DoE policy and the IEB policy views, and practices, of cognitive demand are different [Chapter 4; Section 6.2.2.2]).

Therefore, how much equivalence is ‘enough’ between examination question papers? In South Africa, the statistical adjustment of raw scores is claimed to achieve equivalence across years, subjects and examining bodies (Fatti, 2006). Post-2000, it was also assumed that equivalence or similarity of examinations would result from the standardization of the CBS policy by all state-governed examining bodies, and one common SC Biology examination, and that this standardization was the result of a common understanding of what was important for examination. But what common understanding? Is that common understanding based on explicit policy, or is it made up only implicitly and intangibly as the examiners set examinations? If the policy is explicit, is the policy theoretically and empirically supported? Unfortunately, the archival records about SC Biology examinations, including the names and affiliations of persons responsible for policy decisions, are poor. Intelligent educational policy decisions should be made on the critical examination of current and past educational policies and practices (Dewey, 1938). Without information about many of the educational decisions that

226 Sections of marking memoranda from Biology school leaving examinations of different countries were given in Gandal (1994). Examinations in England and Wales carried approximately one mark per fact; the French indicated marks on the question paper but not on the memorandum, and more than one fact per mark was required in the answers and the Germans indicated mark allocations on neither the question paper nor the accompanying memorandum. The practices of France and Germany indicated that the meaning or context of a fact was considered important in allocating marks.

227 Nitko and Brookhart (2007, p.119) suggest estimates of the approximate time requirements for different types of questions for students in middle and senior high school which might be useful to this conversation.

informed the SC Biology examination practices (and the current NSC practices), it was difficult to understand the philosophy, if any, that informed the various policy decisions that were made during the period of this study.

6.5 Optimal length of examinations

During the exploratory phases of this study, the author performed correlation analyses to attempt to make sense of possible relationships between the performance data. A secondary objective was to measure the constancy of relationships between marks which students achieved. Some of these associations were discussed above, for example, relationships between the marks obtained in different topics or different performance expectations. During the analyses of the question papers, the author was struck by the ‘similarity’ between the different sections and the five questions which comprised each question paper (Section 6.1.1.1). Correlations between each of the questions and the total mark achieved by students for each of the eight 2005 and 2006 SC Biology examinations were surprisingly high (Tables 6.56 & 6.57), as were the correlations between the same students performance on Paper 1 and Paper 2 of the same examinations (Table 6.59). The query arose, if only an aggregate mark was needed to classify students “how long should a SC Biology question paper be?”. While this question was not one of the research sub-questions to be addressed by this thesis, the analyses from this thesis suggest that it is a question that ought to be addressed in the future. The rationale for this suggestion is as follows.

Assessment is defined as a “process for obtaining information that is used for making decisions about students” (Nitko & Brookhart, 2007, p. 508). The SC/NSC examinations are held so that information about students can be collected to provide valid evidence on which to base inferential judgments about students’ abilities. Accordingly, students writing the SC/NSC examinations are classified according to certain levels of mastery (the symbols A to H), and based on this level of mastery whether they qualified for an ME or not. Longer examinations should provide more or better information about students than shorter examination (Figure 6.38 [B]), to justify the increased costs associated with longer examinations. The duration of all the SC Biology examinations prior to 2001 were three hours (Y1 on Figure 6.38 [C]). Post-2001 the duration of the national DoE question papers increased to four hours (Y2 on Figure 6.38 [C]). Currently the NSC Life Sciences questions papers are five hours long (Y3 on Figure 6.38C). Increasing the duration of these examinations must have increased the budget accordingly.

Table 6.56 Correlations between marks obtained in each of the five questions comprising the 2005 SC Biology examinations.

	Paper 1	Question 1	Question 2	Question 3	Question 4	Question 5	Total mark
HG (n = 1001)	**Question 1**	1.00	*0.85*	*0.87*	*0.81*	*0.78*	***0.95***
	Question 2		1.00	*0.85*	*0.80*	*0.80*	***0.93***
	Question 3			1.00	*0.81*	*0.80*	***0.94***
	Question 4				1.00	*0.78*	***0.79***
	Question 5					1.00	***0.90***
	Total mark						1.00
SG (n = 982)	**Question 1**	1.00	*0.85*	*0.81*	*0.86*	*0.73*	***0.96***
	Question 2		1.00	*0.79*	*0.83*	*0.66*	***0.92***
	Question 3			1.00	*0.77*	*0.70*	***0.89***
	Question 4				1.00	*0.67*	***0.92***
	Question 5					1.00	***0.81***
	Total mark						1.00

	Paper 2	Question 1	Question 1	Question 1	Question 1	Question 1	Total mark
HG (n = 1004)	**Question 1**	1.00	*0.83*	*0.83*	*0.86*	*0.84*	***0.96***
	Question 2		1.00	*0.80*	*0.80*	*0.78*	***0.91***
	Question 3			1.00	*0.78*	*0.80*	***0.91***
	Question 4				1.00	*0.83*	***0.92***
	Question 5					1.00	***0.92***
	Total mark						1.00
SG (n = 854)	**Question 1**	1.00	*0.72*	*0.72*	*0.79*	*0.77*	***0.94***
	Question 2		1.00	*0.62*	*0.68*	*0.68*	***0.83***
	Question 3			1.00	*0.71*	*0.69*	***0.82***
	Question 4				1.00	*0.78*	***0.90***
	Question 5					1.00	***0.88***
	Total mark						1.00

Table 6.57 Correlations between marks obtained in each of the five questions comprising the 2006 SC Biology examinations.

	Paper 1	Question 1	Question 2	Question 3	Question 4	Question 5	Total mark
HG (n = 975)	Question 1	1.00	*0.84*	*0.74*	*0.80*	*0.83*	***0.94***
	Question 2		1.00	*0.71*	*0.77*	*0.79*	***0.90***
	Question 3			1.00	*0.75*	*0.74*	***0.85***
	Question 4				1.00	*0.79*	***0.90***
	Question 5					1.00	***0.92***
	Total mark						1.00
SG (n =881)	Question 1	1.00	*0.83*	*0.77*	*0.78*	*0.70*	***0.95***
	Question 2		1.00	*0.77*	*0.79*	*0.67*	***0.92***
	Question 3			1.00	*0.72*	*0.63*	***0.87***
	Question 4				1.00	*0.64*	***0.88***
	Question 5					1.00	***0.80***
	Total mark						1.00

	Paper 2	Question 1	Question 1	Question 1	Question 1	Question 1	Total mark
HG (n = 975)	Question 1	1.00	*0.87*	*0.87*	*0.84*	*0.83*	***0.96***
	Question 2		1.00	*0.86*	*0.81*	*0.80*	***0.93***
	Question 3			1.00	*0.83*	*0.80*	***0.94***
	Question 4				1.00	*0.80*	***0.91***
	Question 5					1.00	***0.90***
	Total mark						1.00
SG (n = 881)	Question 1	1.00	*0.78*	*0.78*	*0.78*	*0.82*	***0.95***
	Question 2		1.00	*0.77*	*0.71*	*0.75*	***0.88***
	Question 3			1.00	*0.73*	*0.75*	***0.88***
	Question 4				1.00	*0.73*	***0.87***
	Question 5					1.00	***0.90***
	Total mark						1.00

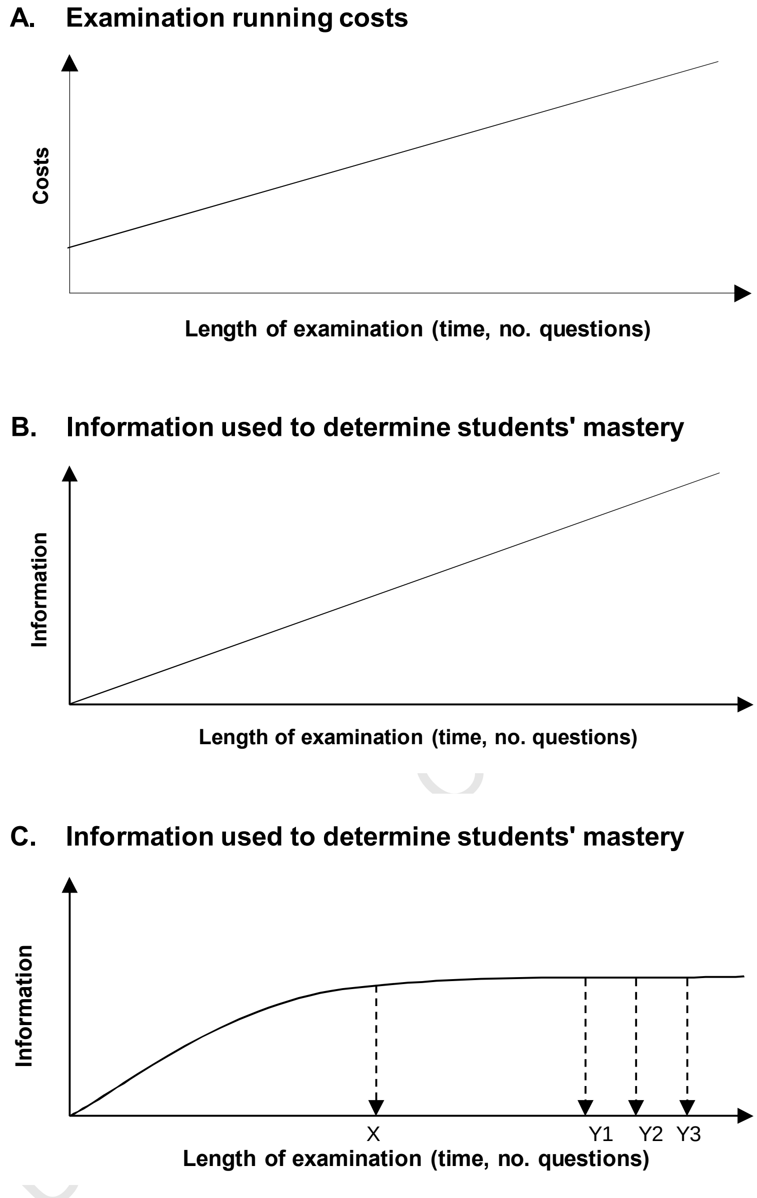


Figure 6.38 Examination inputs and information about students' abilities.

After the initial costs of setting up examinations, the costs of examinations increase with the lengths of examinations (Figure 6.38 [A]) and with an increase in the proportion of free-response answers (Britton et al., 1996).[228] The only publically available figures of the costs of the SC examinations are the figures for the 1999 SC examinations (Asmal, 1999) (Rufus Poliah, personal communication, 8 August 2011). In 1999, the total budget for the administration of the SC examinations for all nine provincial education departments amounted to R303 million[229] – an amount that was considered to be inadequate at the time (Asmal, 1999). The introduction of national DoE administered examinations in 2001 increased this amount considerably (Rufus Poliah, personal communication, 8 August 2011).

The findings a from this study suggest that the amount of information on which the SC Biology examination gathered about students did not rise linearly but instead tapered off much like in Figure 6.38 (C). For example, where r =0.80 then r^2= 0.64, (Table 6.56) about 64% of the squared variation in marks (the response variable) was explained. In the case of 225 HG Paper 1, Question 1 explains 95% of the variance of the total mark (r^2 =0.95) and Questions 2 to 5 do not necessarily provide more information to evaluators about students.

Long ago, Scriven (1974, quoted by Shephard, 2010) stressed the importance of studying the side-effects, the cost and cost-effectiveness of assessment tools, and argued that due consideration be given to other competing assessment tools. More recently, Cizek et al. (2008) and Cizek et al. (2010) demonstrated insufficient defensible validation procedures concerned with test development and test use. In South Africa, a competing assessment tool might be a SC Biology examination with a different design and length. What is the optimal length of a SC Biology examination (X in Figure 6.38 [C])? A great deal of money is spent annually on SC/NSC examinations. If only one mark, the aggregate examination mark, is deemed necessary, irrespective of the meaning of that

228 "When thinking about the cost of assessments, people too often confuse expenditures with cost. Expenditures refer to the dollar amount a state spends on assessment. Costs include both expenditures and the opportunity costs of a particular decision in terms of other valued outcomes. Furthermore, a full analysis of costs should include an estimate of benefits associated with investments. The single dollar figure associated with spending on tests does not capture these trade-offs in the overall education funding system" (Darling-Hammond & Pecheone, 2010, p. 44).

229 R303 million or US$ 49.5 million (R6.12 = US$ 1 [Anon., 1999)]) for 553 299 full-time candidates who wrote a minimum of six subjects and 239 007 part-time students who wrote one or two subjects (Asmal, 1999). Britton and Raizen (1996) provided costs associated the mathematics and science school leaving examinations of seven countries. Darling-Hammond et al. (2010) provided a range of costs associated with large-scale assessment. The range was attributed to the relative proportions of selected-response and free-response questions; the levels of routine within the examination system and the expertise of the markers.

mark, to certify students, then shorter examinations will possibly suffice, and are cheaper to administer and mark, might therefore be more cost effective. These examinations would need to be structurally similar to the longer versions but test the same spectrum of HOCS. Or, the larger sums of money might be better used to develop more sophisticated examinations that would clarify more precisely the meaning of marks that students achieve. This alternative strategy is especially important given the low number of students who passed the SC Biology examinations on both HG and SG (Chapter 2, Table 2.11; Figure 6.37). More explicit information would inform teachers better about exactly what their students need to know and be able to do in order to be successful in the SC Biology examinations.

This question of shorter papers needs further statistical analyses, such as examining how the rank order of students change if, for example, only the marks for Question 1 were used. Such analyses are beyond the scope of this thesis, but the author believes that such analyses should be explored. How might examinations be differently set so that we obtain more meaningful and consistent information about what students know about Biology from high school leaving examinations? Similarly, does separating question papers into different sections, as was done in the SC Biology examinations investigated in this thesis, add value to the information obtained about students?

6.6 Chapter summary

Figure 6.39 describes the complexity of this chapter as it marshals the empirical evidence to directly address, the as yet unanswered research sub-questions 2 to 7 and the research question (Chapter 1). This figure also serves to demonstrate the particular methodology developed in this thesis to conceptualize (Chapter 3) and operationalize (Chapters 4 and 5) standards of the South African SC Biology examinations. In this thesis the focus on validation was interpreted as gathering evidence to investigate the intended meaning of SC Biology examination scores rather seeking evidence to justify the use of SC Biology examinations (Chapter 3). Validity evidence was used to generate the content standards in the SC Biology examinations from 1994 to 2007, and the structural aspects of examinations known to affect student performance therein.

Articulation of the content standards according to an explicit framework developed in this study (Chapter 3) allowed comparisons to be made between years. There was much variation in content standards between the 14 years. For two of these years, 2005 and 2006, the performance standards were explicated for eight examination question papers. The performance standards together with the content standards, bring meaning to the aggregate mark that is used to certify students.

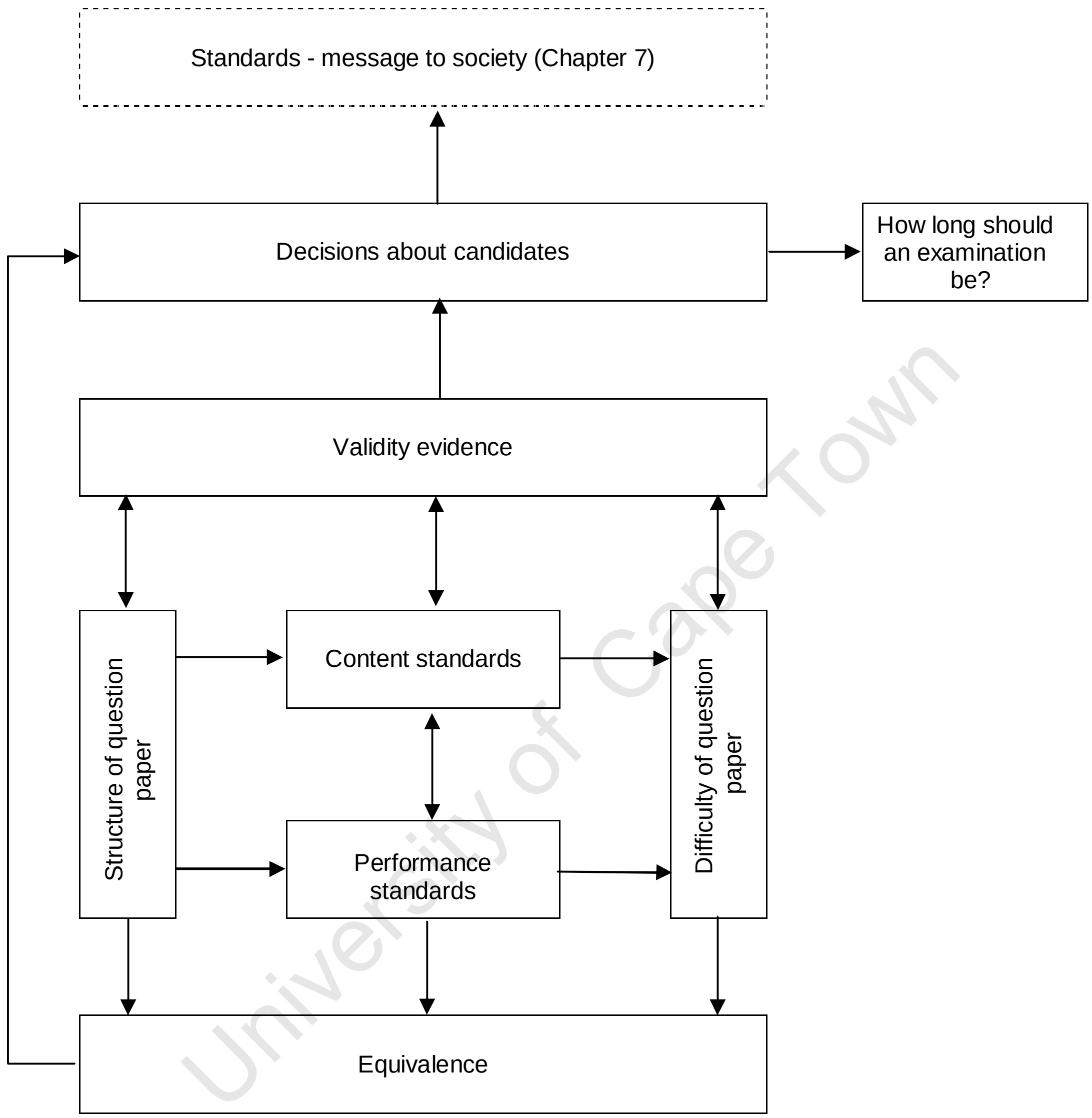


Figure 6.39 Relationships between the different sections of this chapter and Chapter 7.

The impact of the structure of the question papers on student performance was explored, as were the effects of the structure of the question papers, the content standards and performance standards on the difficulty of these eight question papers. These elements varied between comparable question papers in the two years. Performance standards derive their meaning from the content standards on which they are based (Chapter 3). The meanings generated for each of the comparable performance standards were different for each of the two years for which SC Biology examinations scripts were analyzed in this study, because the content standards from which each comparable pair of performance standards was generated was different. Before this study the performance standards in these examinations were arguably meaningless in that they were represented by performance level labels or symbols, with no performance level descriptions. The variability in performance standards impacted on the assumed equivalence of the examinations. Equivalence was a pre-requisite condition for making fair and valid decisions about how students were certified in both years. Therefore SC Senior validation processes cannot focus only on the analyses of examination question papers as is currently the case in South Africa. Validation processes need to included analyses of both a question paper(s) and student answer scripts to determine the consistency, or lack thereof, in the meaning of performance standards.

Educational standards indicate to the society in which they operate what students need to know and be able to do, and how well they need to know and do. Assessments should be used to communicate expectations of the science education system to those concerned with science education (Chapter 3). One message conveyed to society, by SC Biology examinations which formed the basis of this study, was that only one composite mark sans meaning was important in the examination, and an assumption that the one composite mark carries the same implicit but undisclosed degree of meaningfulness between years. The results reported in this chapter are that the meaning, and therefore the value, of an aggregate mark in the 2005 SC Biology examination was different to that of the same aggregate mark in comparable the 2006 SC Biology examination. In addition, explication of these content standards and performance standards of the SC Biology examinations analyzed here provides the empirical evidence for South Africans, especially stakeholders in the SC, to decide if these standards match their intended purpose, and indeed if the SC accurately reflects what is expected of students leaving high school with a certification in SC Biology.

Central to this chapter was the notion that validity evidence brings meaning to scores of students in examinations. Throughout this chapter, interrogation of the validity evidence collected in this study to make meaning of student scores, raised a number of about aspects of both the policies which guided the SC Biology examination processes, and therefore the SC Biology practices. Questions such as these could inform future policies, and therefore the practices, with respect to how Biology is assessed in South African schools, and especially how validation of future NSC Life Sciences examinations is conducted. One incidental question about the optimal length of the SC Biology examinations was addressed in this chapter. If the NSC, which has replaced the SC, continues to require only an aggregate mark, it is proposed that the duration of examinations might not be optimal given the costs of the SC, and now the NSC examinations. The costs might better be used to improve our understanding of the meaning of the aggregate mark and how we might differentiate better students with different levels of Biology mastery. In the final chapter, Chapter 7, the remaining questions generated in this chapter are formulated into a set of recommendations from this study, and a series of potential research questions for future studies.

CHAPTER 7

CONCLUDING CHAPTER

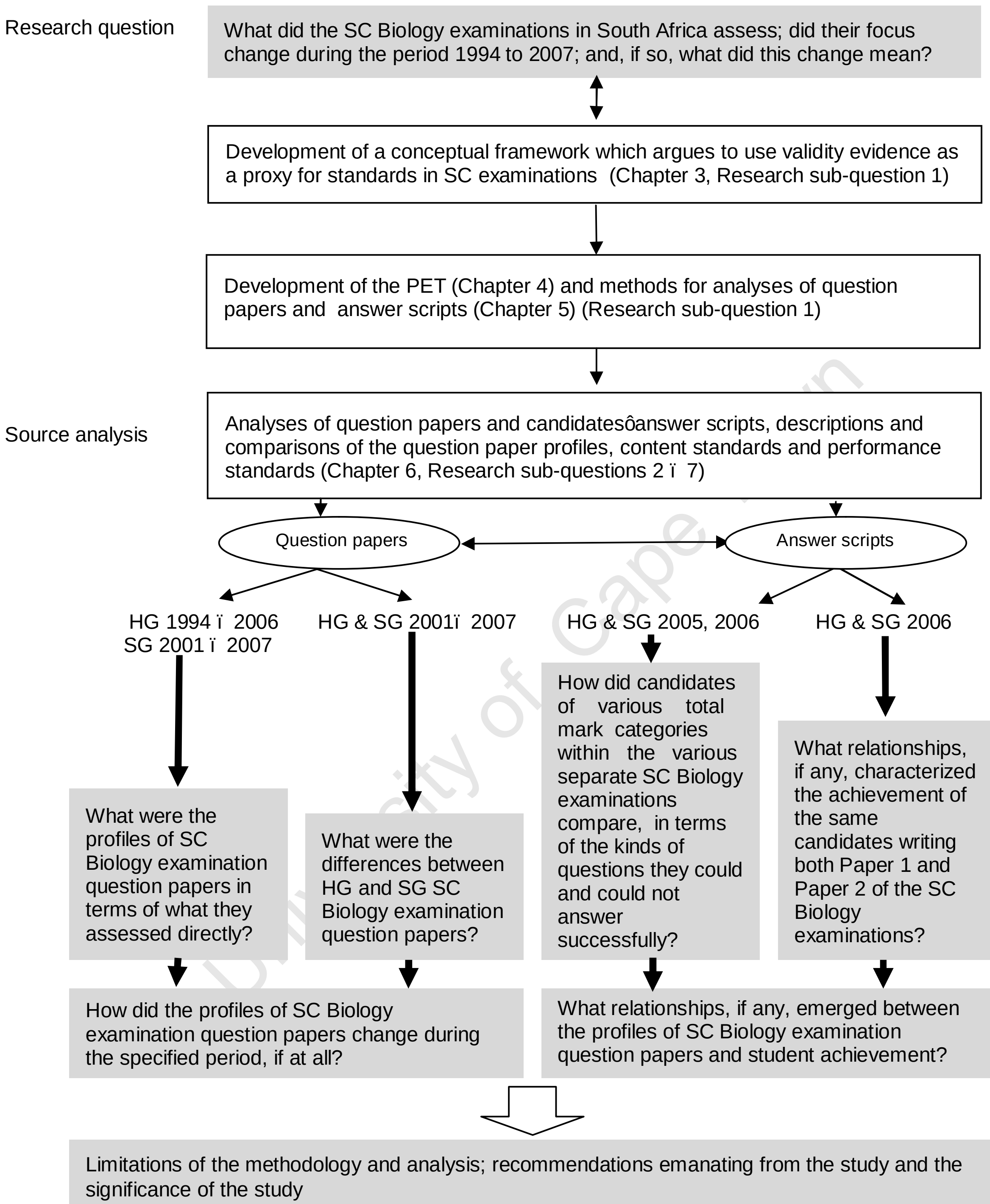


Figure 7.1 The structure of this enquiry into SC Biology examinations and student performances. The research question and the research sub-questions (2 to 7) are addressed in this chapter, together with a discussion of the limitations, recommendations emanating from the study and the significance of the study.

This study seeks to understand what the South African SC Biology examinations assessed during a fourteen-year period of educational transition in the country with the intention of developing a methodology by which these examinations could be both retrospectively understood and prospectively practiced (Chapter 1). The SC examinations were embedded in particular contexts which were quite different from the contexts in which high school leaving examinations operated in any other country about which the author could find comparable information (Chapter 2).

Given the lack of explicit standards about South African SC examinations in general, and Biology examinations specifically, a conceptual framework was built to generate the standards inherent in an extensive set of SC Biology examination question papers and candidates answer scripts (Chapter 3) (Figure 7.1). Such elicited standards were then used to interrogate the focus of these examinations. Two types of educational standards were recognized in this conceptual framework, that is, content standards and performance standards. Content standards were defined as what students should know (topics) and be able to do with what they know (cognitive demand, called 'performance expectations' in this study) (Chapter 3). Performance standards describe different levels of students' mastery of the content standards (Chapter 3). The literature showed that one of the components of content standards, cognitive demand, is problematic to measure consistently and so an instrument—the Performance Expectations Taxonomy (PET)—was developed and validated to measure cognitive demand based on the performance expectations of questions (Chapter 4) (Figure 7.1). In Chapter 5, the operationalization of the conceptual framework to elicit the standards of the SC examinations was completed (Figure 7.1). Both the development of the PET and the methods used to explicate the content standards and the performance standards were rooted in South African SC examination policies and practice described in Chapter 2. Therefore, the composite research sub-question 1, *What are educational standards, and how might they be used to describe and compare SC Biology examination question papers and candidates' answer scripts?* was answered in Chapters 3, 4 and 5.

The conceptual framework developed and operationalized in response to research sub-question 1 used validity evidence to generate the content standards and the performance standards inherent in the SC Biology examinations investigated in this study. For the purposes of this thesis, validity evidence is what supports, or does not support, inferences made about students from the examinations scores achieved in these examinations. The results of the relevant SC examinations were certified, which implied that the examinations were deemed valid as a source of inferences, that is, each of the examinations was of an acceptable standard. Therefore, the examination, in the absence of explicit standards, conveys the actual standards, and we can explicate the implicit standards via validity evidence.

The inferences associated with the SC Biology examinations were, first, that students were reliably classified/sorted according to different levels of competence (i.e., HG or SG; categories of achievement A to H). Second, that students, according to their classification were certified as competent, or not competent, to be considered as citizens of the world and/or as qualified to pursue further study. The second inference depends on the first inference. This thesis is only directly concerned, via the research sub-questions 2 to 7, with the first inference.

The empirical evidence used to answer the remaining research sub-questions 2 to 7 is drawn predominately from Chapter 6. In Chapter 6, the content standards, performance standards and the structural components of examinations which are known to affect these standards, were described and analyzed for possible patterns, and discussed in the light of the policy documents associated with the examinations (Chapter 2). Given the extensive cross-referencing in Chapter 6 between the results of the analyses of the question paper and answer scripts, the author has deliberately chosen, for clarity, to minimize the cross-referencing when addressing the research sub-questions. The answers to all eight research sub-questions, and the results of the analyses of the question papers and answer scripts, are used in combination to address the overall research question, the focus of this thesis, in the next section of this chapter.

The limitations of this study to the SC level of schooling in South Africa, for a single SC subject Biology within the period 1994 to 2007, arose in part from both particularities of the South African context and the incomplete availability of the necessary data sources about the SC examinations to answer the research question (Chapter 2). A short discussion of these limitations is followed by a synopsis of the significance of this study. This synopsis precedes the recommendations about the South African SC examinations which emanated from this thesis and potential research questions for future studies. The final section of this chapter, and the last part of the thesis, are the author's concluding remarks about the thesis which link to the background and rationale given for the study in Chapter 1.

7.1 Providing answers to the research sub-questions

Education standards have various meanings in different contexts but provided they can be explicitly articulated they can provide a consistent framework that can be used to understand various parts of education systems, to make comparisons between specified parts of education systems, and to make comparisons between distinct features of any single part of an education system over a period of time. The answers to research sub-question 1 (Chapters 3, 4 and 5) provided the methodology to explicate the content standards and the performance standards of SC Biology examinations.

In the conceptual framework, performance standards acquired meaning in each examination only in relation to content standards of that examination. Therefore, if the content standards varied between different examinations, so might the meaning of any particular performance standards in those examinations. In addition, specific structural elements of examinations were deemed to influence both the content standards and the performance standards, because they influenced student performance in each examination. The descriptions and comparisons of the structural characteristics, the content standards and the performance standards generated by the analyses of the SC question papers and sample of candidates' answer scripts presented in Chapter 6 are now used to provide the empirical evidence to answer the research sub-questions 2 to 7 below.

7.1.1 Research sub-question 2: What were the profiles of SC Biology examination question papers in terms of what they assessed directly? (1994 to 2007)

This question comments only generally about the SC Biology question paper profiles. Descriptive details of the profiles of the question papers are given in Chapter 6 and details concerning differences between profiles over time, and between HG and SG, are addressed in subsequent research sub-questions.

7.1.1.1 *Structural characteristics of question papers*

All the HG questions papers written during this time period included Sections A, B and C as required by the CBS. However, because various questions papers within each of the years had differing counts of total marks, that is, between 300 and 400, the marks for each section varied, as did the relative weightings of each section across the question papers. The total marks per question paper were not dependent on the time allocated to each question paper.

Section A of all the question papers analyzed in this study comprised various combinations of different types of questions, in contrasting proportions. Similar types of questions were rewarded by varying numbers of marks per question. Examples of section A type questions were MCQs; matching columns, diagrams or labels to descriptions or parts of drawings; matching items and statements; providing one term for a description; providing missing words or labels and some short questions. The IEB Section A questions covered a narrower range of question types than did those of the other government-regulated examining bodies.

The MCQs carried varying numbers of distracters across different question papers and most MCQs tested lower cognitive levels, despite the CBS requirement that they should test both recall and higher intellectual abilities. Some MCQs were more complex than other MCQs in the way that the questions were phrased. There was no consistent mark allocation that matched either the cognitive complexity or complexity in the wording of the question. Similarly, no logical pattern between the marks associated with other kinds of Section A questions was found within and between different question papers. Short questions which often appeared in the Section A were no different to the short questions which appeared in Section B of the same question papers. This pattern suggests that short questions were included in Section A simply to ensure sufficient marks to satisfy the policy requirement for that section, rather than for some educational reason.

A variety of different Section B short questions were observed. In some questions students could select the correct answer while others required students to construct their answer. Choices in Section C questions were generally between two essays, two data response questions, or between a data response question and an essay question. These choices were all within the CBS and the modified CBS requirements. Exceptions in Section C were compulsory or optional MCQs replacing a data response or an essay question and occasions when each of the two alternate questions was part data-response and part essay. MCQs, data-response, and essay questions usually required students to function at different cognitive levels, yet mark allocations between these types of questions did not recognize this difference. Similarly, no distinction could be ascertained in the reporting of results, or the standardization of the marks between students in the same examination who chose different types of questions options, because only the aggregate is reported. Without documentation to support the educational rationale, if any, which resulted in offering particular choices for questions, one can only speculate about the reasons for offering those choices. It may have been that because the effect of type of question on student performance was not known, the equivalence of questions was assumed, or simply that a choice of question offered what was perceived as an easier option and a more difficult option, presumably to serve all students. The IEB SG examinations offered choices of Section C essays, which were shorter than those of the IEB HG examinations – this structure was outside of the CBS policy but examiners found that Section C allowed them to better distinguish between the mastery levels of IEB SG students (E. Nel, personal communication, August 11, 2009). DoE SG question papers had no essay to differentiate between students.

Questions varied in length between one sentence and three sentences long. One-sentence long questions with free-response answers dominated both the HG and SG question papers. Most HG and SG question papers included non-text components in the questions and, but in varying proportions. Diagrams were the predominantly used kind of non-text and included drawings, flowcharts, and photographs as part of questions.

Answers required for free-response questions varied between one or two terms, one to three sentences, and more than three sentences long. Most question papers required students to produce diagrams, and/or tables and/or graphs as part of their answers. There was no consistency observed in how marks were allocated to non-text answers. The diagrams students were required to draw were most often sections through biological structures. When tables were required in an answer, students generally had to contrast structures or processes or sometimes tabulate the results expected from an experiment. A range of different types of graphs such as line graphs, bar charts, histograms and pie charts were required in the answers to some questions. Given the range of structural features associated with various question papers, and that contrasting structural features can influence student performance differently, the following question is posited for future research: *How do the structural features of examinations impact the equivalence of examinations? Could examinations be structured differently to maximize the evidence that they generate about student learning?*

Generally, there was no observable consistency in how marks were allocated to questions. Hence the following questions are posited for future research: *In what ways can the value of a mark in SC Biology examinations be described? How should marks be allocated in the NSC Life Sciences examinations* (which have replaced the SC Biology examinations)?

7.1.1.2 *Content standards of questions papers*

A range of topics could have been examined in the SC Biology examinations (Chapter 2) and, in addition, the CBS stipulated that concepts from previous years could also have been examined. Topics stipulated to be learned in the Grade 12 year were biochemistry, which included enzymes and co-enzymes; angiosperm physiology, which included plant water relations, growth and development and photosynthesis; cellular respiration; aspects of human physiology, which included nutrition, gaseous exchange, excretion and co-ordination; aspects of homeostasis in humans, including osmoregulation in *Amoeba* and *Lumbricus*; and population dynamics. For a period of time the IEB replaced population dynamics with human circulation in Grade 12. Emphases on these topics in teaching and in the SC examinations varied between different examining bodies until 2001 when the modified CBS for government schools provided unsubstantiated weightings for each of these topics. The five most common topics in the SC examinations were biochemistry, plant water relations, human nutrition, excretion and co-ordination, thus giving the question papers a predominantly human biology focus.

Including concepts required to be learned in Grades 10 and 11 in the SC examination at the end of Grade 12 recognized that both vertical and horizontal relationships between the facts, concepts and processes characterize science. This approach allowed examiners the opportunity to test students for the deeper understanding that is now known to be such an important part of learning. Despite this opportunity, only two questions or three scorable events (Chapter 6, Figure 6.5) out of a total of more than 11 000 scorable events involved in this study, tested knowledge that was not explicitly stated for the Grade 12 year. This period of 14 years constitutes an era of lost opportunity to test for an understanding of complex biological knowledge and HOCS, identified internationally as being an important outcome of secondary school science learning.

Neither the CBS nor the modified CBS were clear about what constituted HOCS. The PET category Memorize, defined as the recall or recognition of knowledge, was the most emphasized of the cognitive levels, and the HOCS levels Analyze and Apply were the least emphasized. Despite differences existing between how various instruments might measure cognitive demand (Chapter 4), it is difficult, irrespective of the tool used, to envisage that any of the question papers met the 40% HOCS required for HG question papers and the 20-25% HOCS required for SG question papers, by versions of the CBS. Of particular concern to this author is the startling lack of Analysis questions in this study period. As defined by the PET, Analysis questions require students to show evidence that they were able to make connections, not required by the syllabus or given within the question, using memorized knowledge or routine procedures in familiar contexts. It is these connections that bring integration, understanding and ultimately consequentiality to scientific knowledge. Similarly, being able to apply knowledge to new situations is considered to be a desirable outcome of learning science and a skill vital for success in the 21st century (Chapter 4).

Further, the PET category Explain may have been overemphasized in the results presented in this study, and students may simply have been recalling their answers. Comparisons of the 2001 to 2007 official memoranda from the DoE and the IEB HG and SG examinations indicated that DoE answers were generally marked for facts only, even when questions required students to show evidence of levels of understanding, whereas the IEB generally rewarded students, especially HG students, for showing evidence of understanding if they were required to do so.

There was no consistent relationship between the breadth of knowledge and the depth of knowledge examined in the question papers set by any of the examining bodies across all fourteen years of this study. Generally, the IEB HG question papers examined fewer topics at a greater depth than did the government examining bodies

When content is considered in the way it is conceptualized in this study as a two-dimensional classification of topics and performance expectations, the range of content categories─content standards─varied considerably both within and between years. Given that the examinations are a corpus of possible questions taken from the domain of what was expected to have been learned, the same content coverage in each examination could not have been expected. Not all topics were necessarily explored at all cognitive levels in each examination. The varying emphases on topics and cognitive levels within different examinations, presented challenges for establishing equivalence between different examinations. All the SC examinations analyzed in this study were certified and therefore there was presumably an assumption made that they were equivalent. Hence the following question is posited for future research: *What constituted suitable content equivalence of SC Biology examinations? How equivalent should NSC Life Sciences examinations be with respect to content?*

There was a dearth of HOCS type questions relative to LOCS identified in most of the question papers analyzed in this study. No coherence was required in the answers to questions of the national DoE SC Biology examinations (when longer answers were required by a question, the facts could appear in any order) which suggests that neither HOCS nor coherence in answers were valued in SC Biology examinations. Hence the following questions are posited for future research: *What proportion of an NSC question papers should comprise HOCS questions? How can marks be allocated within any answer memorandum so as to reward the coherence of the student answer?*

7.1.2 Research sub-question 3: What were the differences between HG and SG SC Biology examination question papers? (2001 to 2007)

This research sub-question overlaps somewhat with the research sub-question 4 for the time period 2001 to 2007. Consequently, there will be some overlap between the discussion here and in the following section (i.e., Section 7.1.3). The reason for retaining this research sub-question as a separate entity is to provide some insight into the issue of whether or not HG and SG examinations differed with respect to the kinds of questions that were asked. This exploration is pertinent given that SG examinations were designed to be easier that their HG alternatives, but the “standard of the SG examinations [was] not as clear and acceptable as many [thought]” (Lolwana, 2006, p. 24).

7.1.2.1 *Structural characteristics of question papers*

The CBS and the modified CBS did not give specific guidance as to differences in the type of questions that should be asked in either the HG or the SG question papers, other than that HG question papers should have an additional part─Section C. This section was to have included a data-response type question or an essay question (CBS), or half data-response half mini-essay question (modified CBS) (Chapter 2, Table 2.8). The CBS and the modified CBS also stipulated differences in the proportion of HOCS that should be covered in the HG in comparison to the SG question papers (Chapter 2, Table 2.8).

All the DoE HG examinations included item-statement questions whereas only one of the DoE SG examinations (2002 Paper 1) involved this type of question. Item-statement questions emerge in a number of different forms, but the kind used in the DoE examinations required students to decide which, if any, of two facts or terms or concepts applied to an accompanying statement, for two marks per question. The wording of the statement part of these questions was no more or less complex in either HG or SG questions. The DoE SG examinations included matching columns/diagram questions, which were absent from all but two of the DoE HG examinations (2002, Papers 1 and 2; 2003, Paper 2). For this type of question, students were required to correctly match any one fact or term or concept with one of a range of given descriptions, for two marks per question. This pattern suggests that in using these types of questions, the DoE examiners were sometimes rewarding one fact with one mark in the HG questions and one fact with two marks in the SG examinations. Surprisingly, simple recall questions which required students to fill in single missing words to complete a text or labels missing from diagrams for one mark each, appeared on only the DoE HG question papers and not in any of the DoE SG question papers.

Generally, the IEB HG question papers showed a narrower range of distinct types of questions than did their SG question papers. Similar matching terms/labels to statements questions always carried two marks per question in HG question papers, and only one mark per question in some SG question papers. This contrast suggests a mismatch in the way that the value of a mark was perceived within and between years, and between HG and SG.

Both DoE HG and SG question papers included mainly one sentence long questions, and the predominant form of non-text in questions was diagrams. The IEB HG questions were longer than the IEB SG questions, and diagrams were also the predominant form of non-text used in both HG and SG questions. Both DoE HG and the IEB HG question papers required more free-response answers than did their respective SG question papers, and the HG free-response answers required were longer than the SG analogues. Generally, both the questions asked and the answers required by the IEB HG and SG examinations were longer than those of comparative DoE HG and SG question papers.

In accordance with the CBS and the modified CBS, the DoE SG examinations had no essays. As previously discussed (Chapter 6, Section 6.1.1.2c), the IEB SG essays were different to their HG essays in that they were more descriptive, and less argumentative. In marking, proportionally more marks were available for the recall of factual information for SG than for HG essays. Examiners of the IEB maintained that the SG essay question allowed them to differentiate better between SG students with distinct abilities (James Buchanan [IEB examiner], personal communication, May 23, 2008). The presence of an essay in the IEB SG question papers meant that these question papers required longer answers, in similar proportions to those required by the DoE HG questions over the same time period. Both the DoE and the IEB SG question papers required more one/two word answers than did their comparable SG questions papers.

Overall, similar types of HG questions and SG questions carried similar mark allocations in both HG and SG question papers. Given that the DoE examinations were 400 marks /4 hours (HG) and 300 marks /4hours (SG) and the IEB examinations were 300 or 400marks/3 hours (HG) and 225 or 300 marks/ 3 hours (SG), there was no consistent difference in the value of a mark in terms of the time available to earn each mark between a HG question paper and a SG question paper. Hence the following question is posited for future research: *Could the value of a mark allocated to a particular question be used to differentiate between candidates with different levels of competency in the NSC Life Sciences examinations* (NSC has replaced all SC examinations and have no differentiation into HG and SG)?

7.1.2.2 *Content standards of question papers*

The CBS stipulated the topics to be taught and examined in the HG and SG examinations. The differences between HG and SG with respect to topics to be taught involved how much information was required to be learned about each topic across the two levels. For example, HG students were required to learn, and be assessed on, the functions of different parts of the brain, and SG students were required to learn, and be assessed on the general functions of the brain. From 2001 onwards, the national DoE, through the modified CBS, stipulated the weighting of each topic in each examination (Chapter 2, Table 2.8). Consequently, the topic weightings in the DoE HG and SG examinations were similar (Chapter 6, Table 6.17) across years. The IEB followed the original CBS, with no prescribed topic weightings. Combining the topic weightings for the period 2001 to 2007 suggests that the IEB HG and SG examinations had similar topic emphases which matched the topic emphases of the national DoE HG and SG examinations (Chapter 6, Table 6.17). However, within individual years the topics emphasized in the IEB HG and SG examinations were often quite different (Appendix 6.36, Appendix 6.37). In an examination system such as the SC examination system, which considered specified levels of performance to be equivalent between HG and SG within an examination cycle (Chapter 6 Section 6.4), and between examination cycles (Chapter 2, Section 2.1.1), differences in topic emphases should have been problematic, given that some topics are considered to be more difficult to learn than others. However, the examination system simply avoided this potential challenge to equivalence because the aggregate mark, without attention to how it was composed, was used to classify student performance in these examinations.

When the CBS was modified for the national DoE, the proportion of HOCS did not change for HG, but for SG the original 25% HOCS became 20% − 25% HOCS (Chapter 2, Table 2.8). The modified CBS also indicated that LOCS and HOCS should be determined using a modification of the BTEO. The IEB retained the CBS weightings for HOCS and LOCS and introduced their interpretation of BTEO in 2001 (IEB 2007a). According to the policies prevailing at the time, the SG question papers of both the national DoE and the IEB involved less emphasis on HOCs than did their HG question papers.

Teachers encouraged students to take SC subjects on the SG rather than on the HG level because they felt their students had a better chance of passing the SG SC examinations (Phurutse, 2005). It is therefore surprising that Umalusi (2004) noted its concern that SG examinations showed evidence of declining levels of conceptual challenge, and that SG papers were consistently found to be ‘easy’, considering that they were designed to be easier than HG papers (Lolwana, 2006). Of concern to stakeholders (Reddy, 2006a)was the increasing proportion of Biology SC students, who either wrote the SC Biology examinations on the SG or who failed on the HG, but nonetheless received converted SG passes (Chapter 2, Table 2.10). Therefore, in the period 2001 to 2007, most South African students who passed SC Biology did so by demonstrating low levels of HOCS.

During discussions about the future implementation of the NSC, differentiation into HG and SG was considered unacceptable because of the “experiences that learners have had in its application in this country” (Naidoo 2006, p. 16). Hence the following questions is posited for future research: *How similar, or different, are the questions in NSC Life Sciences examinations* (which have replaced the SC Biology examinations and have no differentiation into HG and SG) *to the HG and SG questions of the pre-2008 years*?

7.1.3 <u>Research sub-question 4:</u> How did the profiles of SC Biology examination question papers change during the specified period, if at all? (1994 to 2007)

This research sub-question concerns change over time and the discussion which follows will deal separately with each of the three eras identified in this study: 1994 and 1995, 1996 to 2000, and 2001 to 2007 (Chapter 2). Of necessity, some of the discussion of this research sub-question will involve repetition of what has already been discussed, but 1994 and 1995 provides the benchmark against which changes in the two subsequent time periods are compared. Because SG question papers were available for the 2001 to 2007 period, they are included for discussion in this research sub-question.

7.1.3.1 *Structural aspects of question papers*

a. ***1994 and 1995 HG***

All SC Biology examinations were three hours long. The total marks for the examinations were 300, 330 or 400, and there was no consistency in the relative marks allocated to each of the Sections A, B and C or to questions of similar type.

Choices in Section C questions were generally between two essays, between two data response questions, or between a data response question and an essay question. These choices were all within the requirements of the CBS and the modified CBS. In 1994 DET students faced only compulsory MCQs in Section C, and in 1995 students could choice between a set of MCQs and an essay question in the same section. The CED 1994 and 1995 Section C questions comprised two alternate questions, each consisting of part data response and part essay.

In both years, one-sentence long questions dominated the examinations of government examining bodies, and free-response answers dominated all the question papers. The IEB question papers had more sentences in their questions. All the HG question papers included non-text components in questions but in varying proportions, within and between years, within examining bodies. Extended answers, that is, more than three sentences long, dominated the answers required by the free-response questions of most examinations. All but two HG question papers (i.e., 1994 HOR and 1995 HOD) required students to articulate some of their answers as non-text. Where students were required to draw graphs, these questions were worth four and six marks each, and two of the graphs required students to draw trends in the data rather than to plot precise points. The significance of mark allocation to graphs in post-2000 practice is discussed below.

b. ***1996 to 2000 HG***

All SC Biology examinations were three hours long. The total marks for the examinations were 300, 320 or 400, and there was no consistency in the relative marks allocated to each of the Sections A, B and C or to questions of similar type. In the 1996 to 2000 HG question papers, similar patterns to those of 1994 and 1995 occurred. Multiple choice questions continued to appear as an option to an essay question in Section C of some examining bodies (i.e., 1996 Northern Cape; 1996 to 2000 Free State).

In 1996 the IEB introduced optional sub-questions in their HG question papers in Section B, and this choice of questions was often from different topics and involved different types of questions. The IEB Section A questions covered a narrower range of question types than did those of the other examining bodies.

In all five years one-sentence long questions with free-response answers dominated the question papers, but this pattern occurred to a lesser extent in the IEB HG question papers where questions were longer. All HG question papers included non-text components in varying proportions in their questions. Extended answers, that is, more than three sentences long, dominated the answers required by the free-response questions of most of the examining bodies. Often the IEB question papers required more extended answers than did those of the government regulated examining bodies. Question papers with the exception of 1996 HG WCED, Northern Cape, Free State; 1997 Free State; 1999 WCED, Northern Province; and 2000 Eastern Cape and Free State, required students to articulate some of their answers as non-text. Where students were required to draw graphs, these questions were worth between two and eight marks each, and most of these questions required students to plot the values associated with data rather than to show trends in the data, in contrast to 1994 and 1995.

c. ***2001-2007 HG and SG***

The IEB HG and SG Biology examinations remained three hours long. The DoE HG and SG examinations separated the topics to be examined into two question papers, each of two hours duration. The total marks for the DoE examinations were 200 marks per HG question paper and 150 marks per SG question papers. The IEB HG question papers were worth 300 or 400 marks each, and the SG questions were worth 225 or 300 marks each.

The relative marks allocated to each of the Sections A, B and C was the same across all years for both the DoE HG and SG question papers, and there was much less variation in the type of questions asked in Section A than was observed in the pre-2001 question papers. The IEB HG and SG examinations also showed consistent relative proportions for Sections A, B and C between years – except in 2002 when both the HG and SG question papers carried more total marks than in the other years.

Multiple choice questions in Section A of both the DoE and the IEB HG and SG question papers were consistently allocated two marks each, irrespective of the complexity of the question. The mark allocation for questions using matching columns or labels in a diagram carried either one or two marks per item in all question papers, with the IEB SG questions showing the greatest inter-question paper variability of this characteristic.

In Section C, the DoE HG examinations included two compulsory 'pseudo' mini-essays, one per question paper, replaced considerably longer choice between one data response question and/or essay question. The modified CBS required that Section C be a mini-essay and at face value, that is, reading the question paper, Section C questions looked like mini-essays. However, scrutiny of the memoranda used to answer these Section C questions showed that approximately 83% of the marks for these questions were awarded to any correct facts irrespective of how the facts were connected. This practice minimizes reward for the logic, organization and coherence which should distinguish essay type answers from other kinds of free-response answers. The DoE SG question papers did not have mini-essays. Contrastingly, Section C in all the IEB HG question papers offered a choice of essays for which the memo allocated 47% of the marks of this question to the manner in which the facts were used to answer the question. The IEB SG question papers offered in Section C, the choice between two shorter essays, and their HG question papers offered a choice between two longer essays. The IEB SG essays rewarded students factual recall in these questions with 65% of the total mark.

In all seven years, one-sentence long questions with free-response answers dominated the HG and SG question papers but to a lesser extent in the IEB question papers where there were considerably longer questions. All HG and SG question papers included non-text components in varying proportions in their questions. Shorter answers from one to three sentences long, dominated the free-response answers required by the DoE HG and SG question papers. Extended answers, that is, more than three sentences long, characterized the answers required by the free-response questions of the IEB HG questions papers, and one or two words and one to three sentence long answers were most prevalent in the IEB SG question papers. All the HG and SG question papers, with the exception of the IEB 2002 SG question papers, to varying degrees, required candidates to present non-text answers. The marks associated with the plotting of graphs in the DoE HG question papers increased from 1.3% of the total marks of the combined Papers 1 and 2 in 2001 to 60% of the total marks of the combined Papers 1 and 2 in 2006 and 2007.

7.1.3.2 *Content standards of question papers*

a. ***1994 and 1995 HG***

Most of the Biology topics were examined in varying proportions by all the examining bodies in the 1994 and 1995 SC Biology examinations. The topics most focused on were biochemistry, plant water relations and aspects of human physiology. The topics least examined were plant hormones and osmoregulation in *Amoeba* and *Lumbricus*.

Cognitive demand was focused on the PET categories Memorize and Explain. All examining bodies, except for the Orange Free State (1994), one of the HOD profiles (1995), and the DET (1995) asked questions which required students to demonstrate their abilities to Perform-Routine-Procedures. Questions involving the HOCS categories Analyze or Apply were used by all examining bodies. The lowest proportions of both of these categories in both years occurred in DET examinations.

Combinations of topic and performance expectations into content categories showed diverse emphases on different combinations of content categories, with some content categories not being examined at all in any of the examination question papers. No topics were examined at all of the different cognitive levels.

There was no consistent relationship between the breadth of knowledge and the depth of knowledge examined by each of the examining bodies in 1994 and 1995.

b. ***1996 to 2000 HG***

Most of topics were examined in varied proportions by all the examining bodies, other than the IEB during the period 1996 to 2000. The topics most focused on were biochemistry, plant water relations and aspects of human physiology. The topics least examined were plant hormones and osmoregulation in *Amoeba* and *Lumbricus*. Generally the IEB HG question papers focused on fewer topics per paper than the government-administered examining bodies. The first year in which the IEB examined human circulation in place of population dynamics, was 1996.

Cognitive demand remained focused on the PET categories Memorize and Explain. All examining bodies, except for the Orange Free State (1996, 1997), asked questions which required students to demonstrate their ability to Perform-Routine-Procedures. In 1996 and 1997 the Orange Free State only included Perform-Routine-Procedures questions as part of optional questions. The HOCS categories analyze and apply were examined by all examining bodies in varying proportions.

As for 1994 and 1996, examination question papers in this second period showed diverse emphases of combinations of topic and performance expectations or content standards. There were always some content categories which were not examined in each of the examination question papers. Very few topics were examined at each of the different cognitive levels.

There was no consistent relationship between the breadth of knowledge and the depth of knowledge examined by each of the examining bodies between 2001 and 2007.

c. ***2001 to 2007 HG and SG***

All topics were emphasized in the DoE HG and SG SC Biology examinations in proportions similar to those stipulated by the modified CBS. The IEB HG and SG examinations, free of the modified CBS prescriptions, continued to examine a variable number of topics each year, and emphasized the topics in different ways. The IEB brought back population dynamics as a topic into the SC to replace human circulation in 2004, and did not examine plant hormones in either the HG or the SG question papers. Osmoregulation in *Amoeba* and *Lumbricus* was removed from the syllabus during this era.

Cognitive demand in both the DoE and the IEB HG question papers, continued to focus on the PET categories Memorize and Explain. The DoE HG question papers covered fewer Analysis questions than the IEB HG question papers. The SG question papers of both examining bodies covered fewer HOCS questions, especially in the Analysis category, than did their HG question papers.

Combinations of topic and performance expectations into content showed diverse emphases. Some content categories were not examined at all.

Breadth of knowledge and depth of knowledge (BOK and DOK) were greater for the DoE HG question papers than they had generally been for government examining bodies in the previous years. As in previous years the IEB HG examinations continued to show no consistent relationship between BOK and DOK. The IEB HG question papers tended to examine a narrower suite of topics, but students had to learn all topics in preparation for the examination as they did not know which topics would be examined or in what proportions each would be examined.

When combining the BOK and DOK data into two periods 1994 to 2000 and 2001 to 2007 for the IEB question papers and the CED-WCED-DoE examining bodies, the IEB HG question papers involved more consistent relationships between BOK and DOK over the entire period 1994 to 2007 than did the CED-WCED-DoE. Generally, the IEB HG question papers tested students over a narrower range of topics at greater depth than was the case with the CED-WCED-DoE HG question papers. The introduction of the modified CBS for DoE students in 2001 resulted in an increased BOK, because for the first time all topics had to be examined in prescribed proportions. A similar increase in DOK occurred because a taxonomy for determining cognitive demand was given for the first time in the modified CBS. Department of Education SG question papers also showed consistent BOK and DOK relationships because of the prescriptions of the modified CBS. The IEB examinations operated under the original CBS for the entire period of this study which accounts for the wider range in the BOK and DOK relationships particularly in the 2001 to 2007 period.

Distinctions between HG and SG question papers were clear within each examining body, and these contrasts were clearer between the IEB HG and SG question papers than between the DoE HG and SG question papers. An exception was the 2002 set of IEB HG and SG question papers. These two papers each had more total marks than corresponding question papers in other years. Because of this change the IEB SG paper in 2002 had the same total marks as the DoE SG examinations that were written in one hour less time. Hence a question is posited for further research: *Can a value be assigned to one mark? Should the value of one mark be tied to the length of an examination? How do the total marks of similar examinations affect their assumed equivalence?*

7.1.4 <u>Research sub-question 5</u>: How did candidates of various total mark categories within the various SC Biology examinations compare, in terms of the kinds of questions they could and could not answer successfully? (2005 and 2006)

This research sub-question concerns the performance standards exhibited in SC Biology examinations over two years, 2005 and 2006. Performance standards derive their meaning from specific content standards (Chapter 3), and the content standards were different for 2005 and 2006 and between HG and SG. The discussion below is organized according to the similarities between the relevant performance standards.

7.1.4.1 *Structural aspects of question papers*

In 2005 and 2006, weaker students, as indicated by their aggregate mark, obtained proportionally more marks for questions which required them to choose the correct answer from answers provided in the question, rather than by constructing their own answers, in both HG and SG cohorts. This pattern might indicate that weaker students guess more or that weaker students are able to answer a question correctly when prompted to do so, or that they are able to recognize some answers more easily that they can recall answers. Unfortunately no data are available to substantiate any view.

There is no consistent pattern in the spectrum of performance standards of students with respect to the length of questions. For example, in 2005 and 2006 HG Paper 1, weaker students do proportionally better in question with fewer sentences. However, in 2005 and 2006 HG Paper 2 this effect was not evident. The converse pattern, that weaker students achieved proportionally more marks from longer questions than from short questions, was noted in the 2005, but not the 2006, SG Paper 1 and Paper 2.

There was no consistent pattern in the use of a particular form of non-text (i.e., table, graph or figure) in questions and associated student performance. In some instances, for example, 2005 SG Paper 1 and 2006 HG Paper 1, weaker students obtained proportionally fewer marks from partially correct answers which included diagrams, than from questions which included graphs or tables, a pattern not observed in the other papers. This lack of consistency is possibly because the use of non-text in a question is invariably linked to the text used in the question, and this relationship was not captured in this study. The lack of a consistent pattern in the use of non-text for an answer and student performance in any of the question papers is confounded by the fact that none of the question papers required the same combinations of non-text answers.

Generally, the relationship between the length of the answer required by questions and student performance, was that weaker students obtained proportionally fewer marks from longer answers, in all eight question papers. A possible explanation might be that weaker students do not have the skills to communicate longer answers, perhaps because many were not writing in their first language. An exception is the 2005 HG Paper 2 where the relative proportions of short answers increased from symbols A to FF and then decreased from FF to H. The 2005 HG Paper 2 was unusual in that it included more HOCS questions than the other 2005 and 2006 HG question papers. It is possible that HOCS and the length of required answers interacted in some way that was not captured by this study. Weaker students may not be guessing the answers in choosing the correct answer. It may be that not having to articulate a free-response answer, helps students to better indicate what they know. If this explanation is valid for South African Biology students, it has implications for the kinds and proportions of choose-the-correct- answer questions versus free-response questions that would have best constituted SC Biology examinations.

Whatever the possible explanation(s), most students who passed SC Biology, passed on the SG and at lower performance levels of SG. This outcome indicates that these students showed little evidence that they could communicate their answers in longer descriptions. Evidence that students can articulate what they have learned (category Explain of the PET) or how they are able to apply their learned knowledge (categories Analyze and Apply of the PET) requires students to be able to articulate answers often longer than three sentences. That these three sentences often did not have to relate to one another, as indicated in the DoE examination memoranda, is a cause of concern, because students are certified as being competent to leave high school. Worldwide HOCS have been identified as important qualities necessary for high school graduates to have. Unfortunately, there is no empirical information available as to exactly what knowledge or skills equip South African SC students to be successful beyond Grade 12.

7.1.4.2 *Content standards of question papers*

Given that performance standards derive their meaning from content standards and that each of the 2005 and 2006 SC Biology examinations tested different subsets of the content standards, the lack of consistency in the relationships between topics examined and performance expectations (i.e., content standards) and student performance is not unexpected. The lack of consistency is exacerbated because the cut-scores which delineated each aggregate symbol remained the same for the 2005 and 2006 examinations. General international practice is that cut-scores are adjusted for each examination, based on student performance, to accommodate differences between the content standards examined (Chapter 3).

The only topic which appeared to be found increasingly difficult by HG and SG learners was population dynamics. Populations dynamics does not directly relate to any of the other topics examined at SC level, but it is related best to concepts studied under the ecology section of the Grade 10 intended curriculum. It might be that weaker students were unable to make the conceptual connections between population dynamics and ecology that were necessary to answer the population dynamics questions, or that teachers failed to link the facts and concepts covered in this section of work to what students had learned in Grade 10. This study has no data to support or refute this possible explanation. In addition, the 2005 HG paper 1 and SG Paper 1 population dynamics questions include the interpretation of experimental results, and such questions, requiring HOCS, were found to be difficult for all students, irrespective of their ability level.

Generally, the proportion of the marks derived from LOCS and HOCS remained similar for all students irrespective of their ability, and resembled the proportions of HOCS to LOCS on each question paper. This pattern means that differing symbols did not capture differential performance in HOCS relative to LOCS. That is, better performing A students did not show a higher proportion of HOCS marks relative to their LOCS marks than did the weaker H students. Exceptions were the proportion of HOCS to LOCS decreased slightly (from symbols A to H) in the 2005 HG Paper 1 and Paper 2, and the proportion of marks achieved from HOCS relative to LOCS increased slightly (from symbols A to H) in the 2006 SG Paper 2.

The aggregate symbols, A to H, denoted the performance standards relative to the content standards examined, and each symbol should have had a performance level descriptions—PLDs—to describe its meaning. Where examinations are considered to be equivalent, common PLDs should apply to the equivalent examinations. In this study PLDs could not be derived to describe the symbols A to H because of differences in the subset of content examined in each comparable examination. It might have been possible to generate PLDs for the examinations analyzed in this study had different cut-scores been created for each comparable examination (i.e., HG 2005 and HG 2006; SG 2005 and SG 2006). This approach was not attempted in this study because of the way the sample of student scripts was selected to represent each symbol. Hence the following questions are posited for future research: *Could the cut-scores in each of the 2005 and 2006 HG and SG SC Biology examinations have been set differently*, *so as to yield consistent PLD*s *between the two years? If the cut-scores between symbols could be set differently for these examinations, how might the cut-score methodology apply to the NSC Life Sciences examinations?*

7.1.5 Research sub-question 6: What relationships, if any, characterized the achievements of the same candidates writing both Paper 1 and Paper 2 of the SC Biology examinations? (2006)

Student results of candidates writing Paper 1 and Paper 2 of the same examination are expected to be similar for the common question paper characteristics, if the results are reliable. In both 2006 examinations, HG and SG students achieved the same or better symbols on Paper 1 than on Paper 2 which suggests that Paper 2 was more difficult than Paper 1. Since the two question papers were introduced in 2001, Paper 2 became known in South Africa as the more difficult of the two question papers because of the combination of topics it examined. The data generated from this study cannot identify the reasons why Paper 2 was more difficult than Paper 1 because it was not only the topics that differed between the two papers, but the structural aspects of the question papers as well.

Associations in the performance of the HG candidates in Paper 1 and in Paper 2 with respect to the particular structural characteristics of the question papers, were similar for choose-the-correct-answer questions and the free response answer questions; and for length of the answer required and levels of cognitive demand. Associations in the performance of the SG candidates in Paper 1 and in Paper 2 were only similar for choose-the-correct-answer questions and the free response answer questions, and the length of the answers required.

Both HG question papers included similar proportions of HOCS and LOCS, and students performed similarly in each of the individual PET categories across both papers. The SG Paper 2 tested more HOCS than did SG Paper 1, and the SG students found the fewer Explain (LOCS) questions on Paper 1, than Explain questions in Paper 2 more difficult. Standard Grade Paper 2 included more Apply questions than were in Paper 1 and students found some of these questions more difficult that the Paper 1 Apply questions.

The lack of clear relationships between the performance of students in Paper 1 and Paper 2 is possibly a result of a number of differences, namely, in the relative structure of the question paper, in the relative cognitive demand of the question paper; and the artificial use of the same cut-scores for each symbol of aggregate performance in each question paper. This obscurity may be compounded by subtle interactions between the structural elements, topics and cognitive demand of a question paper which were not detected in this study; and by the consequences of the fact that the 'value' of each mark awarded differs across question papers. Hence the following questions are posited for future research: *Could different combinations of the structural aspects and the content standards (topics and performance expectations) explain better the relationships, if any, of Paper 1 and Paper 2 for the two years? If so, how might the NSC Paper 1 and Paper 2 be set to maximize the reliability of the results obtained by students writing both question papers? How might the 'value' of a mark be consistently used when setting NSC Life Sciences question papers?*

7.1.6 Research sub-question 7: What relationships, if any, emerged between the profiles of SC Biology examination question papers and student achievement? (2005 and 2006)

Within each question paper, patterns were observed between student performance across the spectrum of performance standards and the variables constituting each question paper profile. Such patterns were discussed above when answering research sub-questions 5 and 6. However, with the exception of two of the variables constituting a question paper's profile (i.e., length of the answers required and whether the student had to construct their own answer), the patterns between each variable and student performance observed in individual question papers, were not consistently observed across all question papers. Reasons for a lack of consistency were discussed in detail above (research sub-questions 5 and 6) and will not be discussed again. It suffices to repeat that performance standards of an examination take on their meaning, or PLDs, from the specific content standards that were examined. If the content standards are different between examinations, the PLDs will be different, unless the cut-sores which define each transition are adjusted to force the PLDs to be the same between those examinations which are deemed equivalent.

Recently Jansen (2012, p.1) suggested that the high pass rates in the current NSC examinations "belie" the conceptual and skill limitations of [South African NSC] school-leavers". Data from this thesis shows that the previous SC Biology examinations were similarly flawed. The 2005 and 2006 SC Biology question papers included fewer HOCS than LOCS questions and most students passed on lower (SG, lower symbols) than on higher levels (HG, higher symbols).

7.2 Providing answers to the research question

What did the SC Biology examinations in South Africa assess; did their focus change during the period 1994 to 2007; and, if so, what did this change mean?

The SC Biology examinations were administered to collect evidence used to make inferences about students' mastery, or lack of mastery, of Biology at the end of Grade 12. The evidence used by the South African education system to make this inference was the total or aggregate mark that a student achieved in a SC Biology examination. Based on an aggregate mark, a student was assigned to a particular level of mastery and was certified accordingly. Therefore, the SC examinations signalled to the South African society the knowledge and skills that were valued, and the required level of mastery of the knowledge and skills that denoted success at this level of education. Several assumptions were made by the people in the education system responsible for the SC examinations. These are: first that comparable question papers were equivalent between years; second, that the aggregate mark somehow encapsulated a student's level of mastery of Biology; and third, that levels of student mastery were consistently certified between years.

The research question central to this thesis was posited to interrogate these assumptions and consisted of three parts. The first concerned describing what biological knowledge and skills were valued in the SC Biology examinations, and how mastery of these elements was recognized. The second part was concerned with whether what was valued and recognized as mastery therein was consistent, or whether it changed, between the SC Biology examinations in different years, from 1994 to 2007. If what was valued and recognized as mastery in the SC Biology examinations changed between years, this then required engagement with the implications of these changes─the third part of the research question Evidence to answer the research question was sought from both examination question papers and answer scripts.

To answer the research question systematically required the formulation of seven research sub-questions specific to the SC Biology examination context. As each research sub-question has already been answered, they will not be discussed in detail again. Instead, the relationship of each sub-question to the main research question will be briefly explained because they all, together with the empirical evidence from Chapter 6, contribute to the answer of the overall research question. Thereafter, the three parts of the research question will be addressed by summarizing and connecting, in bold text, previous detailed discussions which occur elsewhere in this thesis.

The SC was not a part of an explicit standards-based curriculum, so there was little understanding about the relationship between student achievement in the SC examinations, mastery of subject and standards. Answering research sub-question 1 provided the conceptual framework and the methodology to extract standards from SC Biology HG and SG examinations. Content standards, that is, what student should know and be able to do with what they know, were generated from the analysis of question papers. Extracted with the content standards were the structural characteristics of each question paper that were recognized as influencing student performance in the conceptual framework. The structural characteristics together with the content standards of an examination give each examination a unique profile which was used to address research sub-questions 2, 3 and 4. Performance standards define differential levels of performance, that is, mastery, in the specific content standards of an examination, and were extracted from the analysis of candidates' answer scripts after the content standards of the relevant question papers were identified. Research sub-questions 5 to 7 were concerned with describing and comparing the performance standards for four of the SC Biology examinations for which content standards were generated. Research sub-questions 5 to 7 also tested the relationship between content standards and performance standards postulated in Chapter 3.

<u>Part 1 of the research question</u>: *What did the SC Biology examinations in South Africa assess?*

Two policy documents, that is, the CBS (pre-2001) and the modified CBS (2001 to 2007), directed the general composition and structure of the SC Biology examinations (Chapter 2). During the period covered in this study, the government-regulated examining bodies used the CBS and the modified CBS while the IEB used only the CBS. A lack of explicitness in the original CBS policy about the design of the SC Biology examination question papers (Chapters 2, 4, 6) meant that examiners interpreted these policies differently. Consequently, a range of different styles of question papers and questions, with different emphases on topics and different levels of cognitive demand, were evident when viewed through the lens of standards (Chapter 3, sub-research question 1). Detailed descriptions of the standards of the 111 question papers, and the 7 553 candidates' answer scripts, analysed in this study, together with the implications of the similarities and differences in the standards of different question papers, were discussed in Chapter 6 and with answers to the sub-research questions 2 to 7 above. The possibility that each question paper had a multidimensional gestalt, which was not describable in this study, was mooted in Chapter 6.

<u>Part 2 of the research question</u>: *Did their focus change during the period 1994 to 2007?*

Subtle changes in the focus of the question papers were observed between the period 1994 and 1995, and the period 1996 to 2000. Some of these changes were attributed to changes in examining bodies, to new or different examiners and their particular interpretations of the CBS. The most obvious change in the focus of the SC Biology examinations was as the result of the modified CBS, which was used by the DoE examining body from 2001 until 2007. Details of the change in focus of the examination question papers for these three time periods were discussed in Chapter 6, as well as with answers to the sub-research questions 2 to 7 above.

Part 3 of the research question: *If the focus of the SC Biology examinations changed during the period 1994 to 2007, what did this change mean?*

The implications of changes in the SC Biology examinations during the period of this study have already been extensively discussed in Chapter 6, and with answers to the sub-research questions 2 to 7 above. These implications can be summarized as challenges to the perceived equivalence of SC Biology examinations both within and between years; the kinds of evidence that are necessary to make judgements about the equivalence of these examinations; and the lack of examination questions testing students HOCS abilities.

If the focus of SC examinations and associated student performance therein is to be understood, as is required to determine the equivalence between examinations, it is not enough to examine just the content standards inherent in the question papers. Examination questions papers need to be analyzed together with the answer scripts in each examination. The performance standards cannot be assumed to be the same between successive examinations because the subset of content examined differs between each examination, and therefore the content standards differ between examinations. Students are certified between years on the basis of assumed, but not demonstrated, equivalence of performance standards.

Cut-scores in South Africa cannot validly be assumed the same for all subjects because of differences in how the domain of each subject, and therefore the content standards in each subject, are conceptualized. Within one subject, such as Biology, cut-scores cannot be the same across years, because content standards are not the same across years. What is needed are flexible cut-scores which are able to identify defensible performance standards across the spectrum of student performance consistently between years and which take into account different subsets of content standards.

In South Africa, discussions about the standards SC examinations, and the current NSC examinations, have focused on analyses of question papers only, that is, on the content standards only (Chapter 3). This study has shown that consideration of the performance standards of the examinations is vital if we wish to understand student performance and how we might determine and ensure equivalence of student performance within years and between years.

7.3 Limitations of this study

Many of the limitations of this study are a result of the unique South African SC examination context, but some have resulted from the way that the study was designed. These limitations are acknowledged below.

1. Performance expectations were considered to be a broad indicator of cognitive demand. The performance expectations generated for use in this study were derived from an interpretation, and a re-combination of, views of performance offered by various educators in the literature, together with South African syllabus and guideline documents. This study did not employ think-aloud protocol analysis (e.g., Ericsson & Simon, 1998; Lane, 2011, Li, 2002; Smagorinsky, 1998) to confirm whether the performance expectations did, or did not, capture the real thinking processes that students used when they answered examination questions.

2. No official marking memoranda were available for the examination question papers for the period 1994 to 2000. This lack of access to evidence meant that the cognitive demand of some of the pre-2001 examination questions might possibly have been incorrectly classified, irrespective of which taxonomy of cognitive demand had been used.

3. This study assumed that all students who wrote the SC Biology examinations in 2005 and 2006 had been taught the entire intended curriculum and that they had sufficient time to complete the question papers. This assumption may have been incorrect.

4. All candidate answer scripts are destroyed in June of the year following the examination period in which they were written (Chapter 2). Therefore, individual student responses to examination questions could be consulted for only two of the years, 2005 and 2006, covered by this study. This absence of answer scripts from other years means that the relationship between performance standards and the content standards, described in Chapter 3, could not be described for many of the years for which content standards were generated from the examination question papers (1994 to 2004, and 2007).

5. Candidate answer scripts were used from only one examining body, the WCED. The WCED and the performance of its students in the SC examinations is not representative of performance in all government-administered examination bodies in South Africa.

6. The candidate answer scripts were randomly selected, but in different ways for each of the years 2005 and 2006. Because of this difference in sampling, not all of the same analyses of student performance could be undertaken for both years. While this contrast restricted comparisons that could be made between candidates' performance in 2005 and 2006, the 2006 samples provided different and richer answers to some of the research sub-questions. For example, direct comparisons could be made of student performance in Paper 1 and in Paper 2.

7. No personal information was available for the candidates whose answer scripts comprised the samples. Therefore, it was not possible to identify which candidates were being examined in their home language, or which socio-economic group each candidate represented, or if all students had been taught the required content. While this information might have enriched the interpretations of the results of this study, it was not part of the focus of the study.

8. This study assumed that candidates answer scripts were fairly and accurately marked. The marking was not rechecked for accuracy, but the reliability of the marks was contextualized in Chapter 5. This assumption of fairness and accuracy may not have been correct (Chapter 6).

9. The extent of any generalizability of the data obtained from the analysis of available candidate answer scripts to the population of students who wrote the SC Biology examination in 2005 and 2006 could not be established because necessary data about the population were apparently 'lost' in the education system.

10. The statistical adjustments made by Umalusi to the 2005 and 2006 raw scores could not be obtained. No information was available about the school-based assessment mark which contributed to the statistically adjusted scores (the awarded scores) achieved by students for any of the question papers analyzed Therefore only raw scores from the SC examinations were used for the analyses for the years 2005 and 2006.

7.4 Significance of this study

This study contributes to knowledge in four ways which are each described below.

First, while the conceptual framework developed in this thesis (Chapter 3) was conceived to provide a framework to explicate standards from SC Biology examinations, the framework has the potential be used in explicating standards from any assessments which are part of any non-standards-based curriculum. In this way, the conceptual framework enriches how education standards are understood and the role of standards in education systems, by incorporating a South African context. Because the conceptual framework was born out of a reasoned combination of international practices, the South African SC examinations were investigated from an international perspective – hence the recommendations for the South African SC examination practices given in the section below.

Second, the success of the operationalization of the conceptual framework (Chapters 4 & 5) is evidenced by the descriptions of content standards, performance standards and structural aspects of the SC Biology examinations that could be generated in this study (Chapter 6). The conceptual framework enabled detailed comparisons to be made between various SC examinations on the basis of standards (Chapters 6 & 7).

Third, the development of an instrument, the PET (Chapter 4) to determine the cognitive demand of examinations, is the third way in which this study contributes to knowledge. While, many different instruments have been used to determine the cognitive demand of tasks given to students, few of these have been empirically validated, most lack any theoretical or empirical argument about how they function, and consistency of classifications between raters is not well documented (Chapter 4). The PET developed in this study is demonstrably both easy and reliable to use and is an amalgamation of a number of different instruments, used to classify the cognitive demand of assessments, including Bloom's Taxonomy (BTEO), combined with current knowledge about human cognition.

The conceptual framework provided the structure by which to comprehend the unfolding processes of a disparate South African educational history, and permits detailed analysis of standards that were at best only implicit before this study. Therefore, the final contribution of this study to knowledge is that the study generates the first empirical evidence about standards of South African SC Biology examinations using a defensible conceptual framework,

7. 5 Recommendations for practice and future research

Throughout this thesis the author has argued for principled policies and practices with regard to the SC examinations. In consequence, the recommendations emanating from this thesis are therefore recommendations for both policy and practice. The first two recommendations below represent two philosophical arguments made in this thesis. These propositions are first, that validity should be at the heart of SC examination policy and practices; second, that adoption of an explicit standards-based curriculum makes it so much easier to argue for, or against, validity of the examinations. In some countries validity of examinations is a legal requirement─in South Africa it is not. In South Africa, inferences made about SC examination candidates could in the future become a legal challenge, given the high-stakes of these examinations. The third and fourth recommendations concern what stakeholders should do with the standards of the SC Biology examinations generated by this thesis. Recommendations 5, 6 and 7 provide possible ways that the methodology developed for this study might be used in the study of other SC and NSC examinations. The subsequent set of recommendations, that is, recommendations 8 to 13, concern aspects of SC examination validity evidence, and would possibly fall away if recommendation 1 was accepted. The final recommendation, 14, offers a list of potential research questions identified during this research – research questions which could inform both the policies and practices governing the NSC examinations.

As these recommendations apply to future examinations, references are made to only the NSC, and Life Sciences examinations where appropriate, rather than to the SC Biology examinations which were the subject of this study retrospectively.

1. Validity is at the heart of assessment, and arguably at the heart of education. Validity of the NSC Life Sciences examinations should involve a rigorous interrogation of the inferences desired or to be made about any student's learning from his or her performance in a particular examination. These examinations are high-stake events, which means that the equivalence of these examinations within and between years should be explicitly defensible, using validation processes, rather than assumed.

   In a standards-based curriculum (Chapter 3), the standards are made explicit and hence are the framework against which both teaching and assessment can be aligned and compared. The standards define the intended curriculum and therefore provide a working definition of the domain (content standards) that is to be both taught and examined, and thus inform us about how well students have mastered that domain (performance standards). This explicitness means that the standards provide the framework for judgements to be made as to whether an examination, and student performance therein, has met the standards, or not. Prior explication would remove the element of speculation and opinion about the standards of past SC examinations and of the current NSC examinations, and eliminate the need to explicate the standards of the examinations *post hoc*, as was done in this study.

2. In this study, the standards of the SC Biology examinations were deemed to have been accepted by virtue of the fact that students were certified, or not, on the basis of their performance in these examinations. In Chapter 3, the author argued that educational standards are a social construct, therefore, she alone does not have the prerogative to judge whether the standards in these examinations were sufficient, or not sufficient─that is the role of the stakeholders and the community in which the SC examinations operated. The shakeholders now have the empirical data from this study to make informed judgments about the possibly varying standards of past SC Biology examinations.

3. This thesis made available a defensible methodology (Chapters 3, 4 & 5) to extract and explicate standards from SC Biology examinations. By using this methodology, standards may also be elicited from the post-2007 NSC Life Sciences examinations and benchmarked against the results of this study. Stakeholders would then have the evidence to test comparative opinions about standards of the SC and the NSC examinations that continue to circulate.

4. It is recommended that the applicability of the conceptual model developed in this study to the explication of the standards of other previous SC and NSC Biology or Life Sciences examinations, be tested. The generality of the conceptual model and its operationalization warrants its testing by application to the SC and NSC examinations in other subjects.

5. It is recommended that the suitability of the PET to determine the cognitive demand of Biology formative assessment tasks, and the applicability of the PET to other disciplines be investigated.

6. One aspect of the methodology developed in this thesis was the production of content maps to illustrate the content standards in an examination. These content maps are easy and quick to

generate, and to interpret. They therefore deserve a place in the annual analyses of the NSC examinations.

7. The rationale behind policy decisions about the NSC examination practices, including how cognitive demand should be both understood and recognized, ought to be made explicit and documented so that the decisions can be interrogated and not left often open to individual interpretation, by examiners, moderators and the wider public.

8. The NSC examination policies and their rationale need to be archived so that they can be easily accessed by the public and by researchers. During this research, the author found that policy documents have been lost from public record, and the general level of record-keeping about the SC examinations was inconsistent. Researchers and decision-makers need an accurate history that permits comparisons to be made. By being unaware of, ignoring, or misinterpreting our educational history, we are doomed to make the same mistakes repeatedly.

9. Standards of NSC examinations need to be determined from analyses of *both* question papers and candidates' answer scripts. The NSC examinations only examine a particular subset of the content standards or the intended curriculum so each examination has a unique profile. Analyses of candidates' answer scripts together with their question paper yield the performance standards. The performance standards should exhibit the meaning of student performance in an examination.

The unique profile of an examination cannot be reflected by the performance standards if the cut-scores which define each of the discrete performance standards are the same for each examination. Cut-scores need to be re-set for each examination so that the meaning of each performance standard is consistent within and between years in examinations considered to be equivalent.

10. Samples of scripts from the full spectrum of student performance in each NSC examination should be archived to allow comparisons of student performance to be made between years.

11. Questions in NSC examinations that are found to be invalid at the time of marking, should not result in students being allocated the maximum mark(s) for that question. This mark is subsequently included in the composition of the aggregate mark. Doing so inflates students

marks and implies evidence of competencies which some students do not have. Pre-testing all or some of the items used in NSC examinations might be a solution to this risk.

12. Use of an aggregate mark to classify Biology/Life Sciences students, and the use of fixed cut-scores between years, should not make claim to do much more than rank students within a year. If the design of the question papers remains as noted in this study, whatever the merits of the paper for the year in which it is used, there is nothing in the design structure to warrant comparisons.

13. If our focus remains on the aggregate, then we communicate that the aggregate is the only important outcome, and that how we get that aggregate is unimportant. If the examination that provides the aggregate is focused primarily on LOCS, as it was in most of the SC Biology examinations analyzed in this study, then that focus is primarily what teachers will adopt while teaching. If we really want students to be competent at dealing with HOCS we need to include more HOCS questions in examinations. Then teachers will focus more on ensuring that their students practise HOCS, which is really what science requires (Yager, 1993).

14. This study generated a number of possible research questions for future research. These research questions are listed in Appendix 7.1 and have been grouped according to the sequence in which the questions arose out of discussions in this thesis.

7.6 Concluding remarks

Generally, assessment systems are dynamic and vary between countries (Eckstein & Noah, 1993; Britton & Raizen, 1996, Valverde, 2005). This dynamic variability means that, while an understanding of foreign educational systems may help to understand one's own national system (Griffiths & Howson, 1974), research findings from other countries on their assessments, because they are contextual, may or may not be applicable to the South African education system. This study highlighted how the unique the South African SC examinations are in a global context. There have been few published studies in South Africa aimed at understanding the design of SC examination question papers (e.g., Prinsloo, 2004; Umalusi, 2004). Few studies have explored how the questions papers and student performance are linked, and there are no studies describing the relationship between the SC examination Biology question papers and associated student performance.

Large-scale assessments, like examinations, are important for communicating goals for educators and students to pursue. This communication can lead to a positive impact on instruction and learning, and provide “delineation of content standards” (Lane, 2004, p. 12). This study is a first attempt in South Africa to describe the meaning of, and quantify, the competencies that SC Biology examinations assessed in terms of standards, in the period 1994 to 2007. The competencies assessed in the SC Biology examinations indicate, to both the education system and to the public, what competencies a Biology student should have at the end of their secondary schooling and what standards are conveyed by the examinations.

This research differs from other published studies of Biology examinations for several reasons. First, this study develops and describes a specific way of understanding educational standards concerned with Biology, from the international literature. It builds a conceptual framework which uses international standards practice and which modifies the practice for a South African context. The conceptual framework developed in this thesis argues for the use of validity evidence as a proxy for standards, in the absence of explicit standards, and requires the use of both examination question papers and candidate answer scripts as sources of data. Both sources are necessary to describe the relationships between question papers and associated student achievement, in terms of content standards and performance standards. This research also explored the role of standards in an education system where standards are the language which conveys explicitly and consistently to all parts of the education system what students need to be taught and to learn, how students need to be assessed and how student performance is measured.

Second, the conceptual framework required that the cognitive demand component of content standards be reliably and consistently determined in the analyses of examination question papers. Despite the view that HOCS are the high literacy of educational outcomes (Resnick, 1987) and that HOCS have been linked to economic growth and prosperity (Hanushek, 2003), there is no consistent method from the literature specifying how cognitive demand can be defined and measured in teaching and learning. A specific instrument, the PET, was developed and validated to determine the cognitive demand of questions in the SC Biology examination question papers used in this study. The philosophy and design of the PET was born out of the collective strengths and weaknesses of various cognitive demand classificatory tools which have been used in education, especially the BTEO. The strengths of the PET are that each category of cognitive demand is simply defined by the performances expected from students to demonstrate their ability, or lack of ability, to function at that specific level of cognitive demand, and that the PET unambiguously delineates what are LOCS and HOCS, based on an explicit synthesis of various understandings of what constitutes HOCS.

Third, this study documented and measured changes in the structure of Biology SC examination papers during a period of transition in all spheres of South African life, including education. The period of this study preceded the introduction of a new Biology[230] curriculum and therefore provides a benchmark against which the new NSC Life Sciences examinations can be compared.[231] The NSC examinations which were written for the first time in 2008 do not differentiate between HG and SG and there ought to be an understanding about how best to design questions to differentiate accurately between students with different kinds of abilities (Ebel, 1972) within an examination. This study describes historical variations in how Biology SC question papers differentiated, or did not differentiate between these two categories of students. Duncan Hindle, the Director-General of Basic Education in South Africa, stated that "the National Senior Certificate [NSC] results were also new last year, preventing any comparison with previous years" (Hindle, 2009, p. 13) and Eugenie Rabe, Umalusi's Chief Operating Officer, was quoted as saying "[t]he two systems [SC and NSC] cannot be directly compared as they work from two different sets of assumptions" (Govender, 2009, p. 2). The author of this study believes that in order to evaluate the success, or lack of success, of the new NSC policies, comparisons have to be made with the older SC and its policies. This study provides a conceptual framework to guide such comparisons.

Chisholm (2004c) said that until there was public consensus in South Africa concerning the meaning of educational standards, they would remain in the eye of the beholder. This study suggests the conceptual framework and a methodology to measure quantitatively the information or standards implicit in assessments like examination question papers. That is, it permits a meaning of standards in examinations beyond the eye of the subjective beholder. The methodology, adapted or modified, might be applicable to other subjects and/or assessments in general. More importantly, the methodology can inform future policy decisions about what 21st century assessments should be and how assessment systems should function, in order to provide as much data on student achievement as possible.

This study provides some of the kinds of evidence that Towne, Wise and Winter (2004) consider necessary to inform practice and policy in education within the constraints described by Engelhard (2005). Engelhard called for reasoned judgment, and the consideration of multiple sources of information, when making high-stakes educational decisions, especially with regard to assessment. Because this study will contribute towards a temporal understanding of what constitutes evidence of validity and therefore standards of SC Biology examinations, in a South African context, it offers evidence to support the use of examination results in revising curricula, formulating education policy to include validity evidence in assessment practice especially when designing assessments. Only in the context of such information being available about examinations, can those examinations warrant being

230 Biology became known as Life Sciences and was introduced at Grade 10 level in 2006 and examined for the first time at SC level in 2008.

231 The framework developed in this study is being used to compare the NSC Life Sciences examinations for 2008 to 2011, with the analyses of SC Biology examinations done in this study (Crowe, in preparation). The analyses have been excluded from this thesis because the context of the Life Sciences curriculum is different to that of Biology.

considered as an indication of the 'health' of the education system. If "[a] common goal of testing − [is] to find and extend the boundaries of students' knowledge" (Gelman, 2003, p. 75) then what students show that they know via assessments, like the SC examinations, needs to be explicated and examined. This study does explicate the standards of selected SC Biology examinations, and therefore has implications for the teaching and learning of Biology in South Africa.

In some developing countries[232] the absence of system level information about the quality of student learning that can obtained by conducting their own assessments, or by participating in regional or international assessments, has made it difficult to "gauge overall levels of achievement, to assess the relative performance of particular subgroups, and to monitor changes in performance over time …[and] difficult to determine the effectiveness of government policies designed to improve outcomes in these and other areas" (Greaney & Kellaghan, 2008, p. ix-x). In order to improve educational quality, "some reliable measures of current educational quality [and reliable education statistics] are needed" Heyneman (2004, p.448). The conceptual framework developed in this study offers a way to quantify standards in SC Biology examinations, and possibly the SC examinations of other subjects.

During this research, the dearth of public information, including historical and current education statistics, about the South African SC Biology examinations, and indeed the source documents from which these data could be generated, became clear. Based on its findings (intentional and incidental), this study offers suggestions as to how future records and data collection might be operationalized to ensure that the system maximizes the information that question papers and examination scripts can provide towards the understanding and interpretation of student achievement as a measure of educational quality. Because of the centrality of validity to the conceptual framework used in this study, it also suggests the types of validity evidence that are necessary to satisfy some of the legal requirements of high stakes examinations such as the South African SC examinations. The conceptual model of standards described in this thesis is not necessarily a panacea but it is a first stage in a common language by which to understand the SC Biology examinations.

"These are exciting times in the world of educational assessment; days of urgent demands, unprecedented opportunities, and tantalizing challenges[.] Demands for consequential tests in schools and nations, at larger scales and with higher stakes than we have seen before" (Mislevy, 2008, p. 2). It is hoped that the assumptions and findings of this study will in the future be debated, tested and challenged using more sophisticated psychometric analyses in order to refine our current NSC examination practices in all subjects. Most importantly, the author hopes that this study will provide

232 The author catergorizes South Africa as a developing country with respect to education based on the poor performance of its school students in international tests.

some defensible insights into the expression of the enigmatic concept, educational standards, and a methodology to understand them, that will prompt new, and different, questions to be asked by the stakeholders, including the policymakers, of education in South African schools. Such probing questions are urgently needed, given the crisis in South African education, to inform the national conversation and debate about how educational standards can be understood, measured and more importantly consistently practiced at all levels of the South African school curriculum, especially the SC examinations. We as South Africans need to rigorously interrogate our assessment practices and to learn from international assessment practices, while "re[-]imagining how to answer the [educational] challenges of a new century" (Hess, 2010, p. 6). Strengthening the validation processes associated with the NSC examinations is a good place to start—this study provides a methodology to do this using a lens of standards.

REFERENCES

Abedi, J. (2002). Standardized achievement tests and English Language Learners: Psychometric issues. *Educational Assessment, 8*(3), 231-257.

Abedi, J. (2006). Language issues in item development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 377-398). Mahwah, NJ: Lawrence Erlbaum Associates.

Abedi, J., & Gándara, P. (2006). Performance of English Language Learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice, 25*(4), 36-46.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219-234.

ACT. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. Iowa City, IA: Author. Retrieved August 16, 2011, from http://www.act.org/research/policymakers/reports/reading.html

African National Congress. (1994). *The Reconstruction and Development Programme: A policy framework.* Johannesburg, South Africa: Author.

Aikenhead, G. G. (1997). Toward a cross-cultural first nations science and technology curriculum. *Science Education, 81*(2), 217-238.

Aitchison, J. (1986). *The statistical analysis of compositional data.* London: Chapman and Hall.

Alberts, B. (2009). Redefining science education. *Science, 323*, 437.

Alberts, B. (2010). Reframing science standards. *Science, 329*, 491.

Alberts, R. V. J. (2001). Equating exams as a prerequisite for maintaining standards: experience with Dutch centralised secondary examinations. *Assessment in Education: Principles, Policy & Practice, 8*(3), 353 -367.

Allen, D., & Tanner, K. (2002). Approaches to Cell Biology teaching: Questions about questions. *Cell Biology Education, 1*(Fall), 63-67.

Allott, A. (2001). *Biology for the IB Diploma standard and higher level*. Oxford: Oxford University Press.

Amedahe, F. K. (2001). Combining teacher-assessment scores with external-examination scores for certification: the Ghanaian experience. *Educational Measurement: Issues and Practice, 20*(4), 29-30.

American Association for the Advancement of Science. (1989). *Project 2061: Science for all Americans*. Washington, DC: Author.

American Association for the Advancement of Science. (1990). *Science for all Americans.* New York: Oxford University Press.

American Association for the Advancement of Science. (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.

American Association for the Advancement of Science. (2007). *Atlas of Science Literacy – Volume 2.* Washington, DC: Author.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin, 51,* (2, supplement).

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1974). *Standards for educational and psychological tests.* Washington, DC: Author.

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty and student learning. *Education Policy Analysis Archives, 10*(18), (no pp.). Retrieved December 6, 2005, from http://epa.asc/apaa/v10n18.html

Anderson, R. D. (1995). Curriculum reform. *Phi Delta Kappan*, 77(1), 33-36.

Anderson, L. W., Kwathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R. Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's educational objectives* (Abridged Edition). New York: Longman.

Anderson. L. W., & Sosniak, L. A. (Eds.) (1994). *Bloom's Taxonomy: A forty-year retrospective, Ninety-third Yearbook of the National Society for the Study of Education, Part II*. Chicago: University of Chicago Press.

Andrich, D. (2002). A framework relating outcomes based education and the taxonomy of educational objectives. *Studies in Educational Evaluation, 28*(1), 35-59.

Anon. (1999). *Exchange rates: 1999 - South African Rands to 1 USD*. Retrieved March 27, 2012, from http://www.x-rates.com/d/ZAR/USD/hist1999.html

Asmal, K. (1999). *Minister of Education, Prof*[essor] *Kader Asmal's report on education to the President.* Pretoria, South Africa: Department of Education. Retrieved July 5, 2011, from http://www.polity.org.za/polity/govdocs/reports/education/report30-11-99b.html

Assessment & Qualifications Alliance. 2005. *A basic guide to standard setting.* Retrieved June 25, 2007, from http://www.aqa.org.uk/over/pdf/guideltostandardsetting.pdf

Assessment & Qualifications Alliance. 2007. *Setting the standard – How we maintain standards over time.* Retrieved June 25, 2007, from http://www.aqa.org.uk/over/ standards.php

Ausubel, D. P. (1962). A subsumption theory of meaningful verbal learning and retention. *The Journal of General Psychology, 66*(2), 213-24.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view.* New York: Holt, Rinehart & Winston.

Babbie, E., & Mouton, J. (2009). *The practice of social research.* Cape Town, South Africa: Oxford University Press.

Baird, J. (2000). Are examination standards all in the head? Experiments with examiners' judgments of standards in A level examinations. *Research in Education*, *64*, 91-100.

Baird, J., Cresswell, M., & Newton, P. (2000). Would the *real* gold standard please step forward. *Research Papers in Education, 15*(2), 213-229.

Bakker, S., & Wolf, A. (2001). Editorial: Examinations and entry to university: pressure and change in a mass system. *Assessment in Education, 8*(3)*,* 285-290.

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2002). *The analysis and interpretation of multivariate data for social scientists.* Miami, FL: CRC Press.

Baughman, G. D., & Mayrhofer, A. (1965). Leadership Training Project: A final report. *Journal of Secondary Education, 40*, 369-372.

Beach, R. W. (2011). Issues in analyzing alignment of Language Arts Common Core Standards with state standards. *Educational Researcher*, *40*(4), 179-182.

Beck, M. D. (2007). Review and other views: "Alignment" as a psychometric issue. *Applied Measurement in Education. 20*(1), 127-135.

Behr, A. L. (1985). The Senior Certificate Examination as a predictor of university success. *South African Journal of Education, 5*(3), 107-112.

Bernstein, B. (1977). *Class, codes and control, Volume III*: *Towards a theory of educational transmissions*. London: Routledge & Kegan.

Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice, 22*(3), 21-29.

Bhorat, H., & Oosthuizen, (2006). *Final Report III – as part of the project Educational Outcomes in South Africa: A production function approach: Determinants of Grade 12 pass rates*. Cape Town, South Africa: Development Policy Research Unit School of Economics, University of Cape Town.

Biggs, J. B. (1999). Formulating and clarifying curriculum objectives: setting up criterion-referenced objectives. In J. B. Biggs (Ed.), *Teaching for quality learning at university* (pp. 46-49). Buckingham, England: Society for Research into Higher Education and Open University Press. Retrieved February 13, 2007, from http://www.teaching.polyu.edu.hk/Datafiles/R130.pdf

Biology's Big Bang. (2007, June 16). *The Economist, 383,* 13. Retrieved July 24, 2007, from http://web.ebscohost.com/ehost/detail?vid=4&hid=106&sid=032b3821-7387-4a71-ab35-6848694564b%40 sessionmgr106

Bissell, A. N. & Lemons, P. P. (2006). A new method for assessing critical thinking in the classroom. *Bioscience, 56*(1), 66-72.

Black, P. (1996). Commentary Part One: Looking inside the examinations. In E. D. Britton & S. A. Raizen (Eds.), Examining the examinations: An international comparison of science and mathematics examinations for college-bound students (pp. 19-21). Boston: Kluwer Academic Publishers.

Black, P. (1998). *Testing: Friend or foe? Theory and practice of assessment and testing*. London: Falmer Press.

Black, P. (2002). *Report to the Qualifications Development Group Ministry of Education, New Zealand on the proposals for development of the National Certificate of Educational Achievement.* London: Author. Retrieved January 5, 2006, from http://www.minedu.govt.nz/web/downloadable/dl15591_vl/paul-black-report.pdf

Black, P. & Wiliam, D. (2004). *International approaches to science assessment*. Commissioned paper prepared for the National Research Council's Committee on Test design for K-12 Science Achievement, Washington, DC: National Academies Press. Retrieved January 5, 2006, from http://www7.nationalacadamies.org/bota/International%20Perspectives.pdf

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2), 2-5.

Bloom, B. S. (1974). Implications of the IEA Studies for curriculum and instruction. *School Review, 82*(3), 413-435.

Bloom, B. S. (1975). *Evaluation, instruction and policy making*. Seminar Paper 9. Paris: International Institute for Educational Planning.

Bloom. B. S., Engelhart, M. D., Furst, R. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* New York: David McKay.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill Book Company.

Board, C., & Whitney, D. R. (1972). The effect of selected poor item-writing practices on test difficulty, reliability and validity. *Journal of Educational Measurement, 9*(3), 225-233.

Bogan, Y. K. H., & Porter, R. C. (2005). On the ball with higher order thinking. *Teaching PreK-8, 36*(3)*,* 46-47.

Boner, K. (2000). Dominican women − A time to speak. Pietermaritzburg, South Africa: Cluster Publications.

Booker, M. J. (2007). A roof without walls: Benjamin Bloom's Taxonomy and the misdirection of American education. *Academic Questions*, *20*(4): 347–355.

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., & Franić, S. (2009). The end of construct validity. In R.W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 135-170). Charlotte, NC: Information Age Publishing.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (1999). *How people learn*. Washington, DC: National Academies Press.

Brennan, R. L. (2001). Some problems, pitfalls and paradoxes in educational measurement. *Educational Measurement: Issues and Practice, 20*(4), 6-18.

Brennan, R. L. (2006a). Perspectives on the evolution and future of education measurement. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.1-16). Westport, CT: Praeger Publishers.

Brennan, R. L. (Ed.). (2006b). *Educational measurement* (4th ed.). Westport, CT: Praeger Publishers.

Bricker, L. (2002). Accountability and assessments. *Research for Better Schools, 6*(1), 7.

Brindley, G. (Ed.). (2000*). Studies in immigrant English language assessment, Volume 1.* Sydney, Australia: Macquarie University.

Britton, E. [D.] (1996). Study background. In E. D. Britton & S. A. Raizen (Eds.), *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students* (pp. 1-18). Boston: Kluwer Academic Publishers.

Britton, E. [D.], Dossey, J., Eubanks, D., Eubanks, L., Gisselberg, K., Hawkins, S., Raizen, S., & Tamir, P. (1996a). Comparing examinations across subjects and countries. In E. D. Britton & S. A. Raizen (Eds.), *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students (pp. 23-54).* Boston: Kluwer Academic Publishers.

Britton, E, [D.], Hawkins, S., & Gandal, M. (1996b) Comparing examination systems. In E. D. Britton & S. A. Raizen (Eds*.),* *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students* (pp. 201-218). Boston: Kluwer Academic Publishers.

Britton, E. D., & Raizen, S. A. (Eds.) (1996). *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students*. Boston: Kluwer Academic Publishers.

Broadfoot, P. (1979). *Assessment, schools and society*. London: Methuen & Co.

Brown, L. (Ed.). (1993). *The new shorter Oxford English dictionary on historical principles [−] Volume 1 A-M.* Oxford, England: Clarendon Press.

Brown, R. S., & Conley, D. T. (2007). Comparing state high school assessments to standards for success in entry-level university courses. *Educational Assessment, 12*(2), 137-160.

Bruce, M. G. (1991). The standards debate across the Atlantic. *Educational Leadership, 48*(5), 31-33.

Bybee, R. [W.], & Ferrini-Mundy, J. (1997). Guest Editorial. *School Science and Mathematics, 97*(6), 281-282.

Bybee, R. W., Ferrini-Mundy, J., & Loucks-Horsley, S. (1997). National standards in school science and mathematics. *School Science and Mathematics, 97*(6), 325-334.

Bybee, R. W. (2009). *K-12 Engineering Standards: Opportunities and barriers.* A presentation for the Workshop on Standards for K-12 Engineering Education, National Academy of Engineering, Keck Center of the National Academies, Washington, DC. Retrieved June 22, 2010, from http://www.nae.edu/File.aspx?id=15165

Case, B. J., & Zucker, S. (2005a). *Methodologies for alignment of standards and assessments. Policy report*. San Antonio, TX: Harcourt Assessment Inc.

Case, B. J., & Zucker, S. (2005b). *Horizontal and vertical alignment*. Policy report. San Antonio, TX: Harcourt Assessment Inc.

Case, B. J., Jorgensen, M. A., & Zucker, S. (2004). *Alignment of educational assessment. Assessment report*. San Antonio, TX: Harcourt Assessment Inc.

Chalhoub-Deville, M. (2009). Content validity considerations in language testing contexts. In R.W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 39-64). Charlotte, NC: Information Age Publishing.

Champagne, A. B. (1990). *Definition and assessment of the higher-order cognitive skills.* Research Matters – to the Science Teacher, Report No. 9003. Retrieved June 15, 2000, from http://www.narst.org/publications/research/high2.htm

Champagne, A. B., & Klopfer, L. E. (1981). Structuring process skills and the solution of verbal problems involving science concepts. *Science Education, 65*(5), 493-511.

Chang, C. (2010). Does problem-solving = prior knowledge + reasoning skills in Earth Science? An exploratory study. *Research in Science Teaching, 40,* 103-116.

Charoux, M. O. (1993). *The TED Matriculation, Higher Grade, Biology Examination as an evaluator of educational objectives for effective Biology teaching.* Master's dissertation, University of the Witwatersrand, Johannesburg, South Africa.

Christie, P. (1987*). The right to learn.* Johannesburg, South Africa: Ravan Press.

Chisolm, L. (2004a). Matric results 2003 – public debate lacks quality information and analysis. *Human Sciences Research Council Review, 2*(1), 4-5.

Chisolm, L. (2004b). Marking matric: Overview of the debate. *Human Sciences Research Council Review, 2*(4), 6-7.

Chisolm, L. (2004c, February 9). Matric pool redefined. *Financial Mail,* South Africa.

Chomsky, N. (1957). *Syntactic structures.* The Hague, Netherlands: Mouton.

Choppin, B. (1981). Is education getting better? *British Educational Research Journal, 7*(1), 1-15.

Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice, 20*(4), 19-27.

Cizek, G. J. (2006). Standard setting. . In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 225-258). Mahwah, NJ: Lawrence Erlbaum Associates.

Cizek, G. J., Bowen, D., & Church, K. (2010). Sources of validity evidence for educational and psychological tests: A follow-up study. *Educational and Psychological Measurement, 70*(5), 732-743.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting - A guide to establishing and evaluating performance standards on tests.* Thousand Oaks, CA: Sage Publications.

Cizek, G. J., Rosenberg, S. L., & Koons, H. H. (2008).Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement, 68*(3), 397-412.

Cizek, G. J., Webb, L. C., & Kalohn, J. C. (1995). The use of cognitive taxonomies in licensure and certification test development. *Evaluation & the Health Professions, 18*(1), 77-91.

Cobb, P., & Jackson, K. (2011). Assessing the quality of the Common Core Standards for Mathematics. *Educational Researcher*, *40*(4), 183-185.

Collins, A. (1998). National science education standards: A political document. *Journal of Research on Science Teaching, 35(*7), 711-727.

Cohen, A. S., Kane, M. T., & Crooks, T. J. (1999). A generalizable examinee-centered method for setting standards on achievement tests. *Applied Measurement in Education, 12*(3), 343-366.

Cone, J. D., & Foster, S. L. (1991). Training in measurement: Always the bridesmaid. *American Psychologist, 46*(6), 653-654.

Congdon, P. J., & McQueen, J. (2000). The stability of rater severity in large-scale assessment programs. *Journal of Educational Measurement, 37*(2), 163-178.

Conley, D. T. (2003). *Understanding university success.* Eugene, OR: Centre for Education Policy Research, University of Oregon.

Council of Chief State School Officers. (2002). *Models for alignment analysis and assistance to states.* Washington, DC: Author. Retrieved November 15, 2005, from http://www.ccsso.org/contents/pdfs/AlignmentModels.pdf

Council of Chief State School Officers &Wisconsin Center for Educational Research. (2004). *Coding procedures for curriculum content analyses.* Retrieved June 13, 2006, from http://seconline.wceruw.org/Reference/CntCodingProcedures.pdf

Cresswell, M. J. (1986). Examination grades: How many should there be? *British Educational Research Journal, 12(*1), 37-54.

Cresswell, M. J. (1998). What are examination standards? The role of values in large scale assessment. In National Education Authority (Ed.), *The effects and related problems of large scale testing in educational assessment* (pp. 245 - 262). Proceedings of the 22nd International Conference of the IAEA, Beijing, China: Foreign Language Testing and Research Press.

Cresswell, M. J. (2000). The role of public examinations in defining and monitoring standards. *Proceedings of the British Academy, 102,* 69-120.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Laurence Erlbaum Associates.

Cronbach, L. J. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 147-171). Urbana, IL: University of Illinois Press.

Cronbach, L. J.,& Meehl, P. E. (1955). Construct validity: Construct representation in psychological tests. *Psychological Bulletin, 52*(4)*,* 281-302.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

Crowe, A. A. (2000). *A response to a report on the international benchmarking of the 1999 Senior Certificate Biology Examination.* Unpublished mimeograph.

Crowe, A. A. (2004). *A response to a report on an investigation into the standard of the Senior Certificate Biology Examination.* Unpublished mimeograph.

Crowe, A. A. (2007a). *External moderation of Biology Senior Certificate CASS – NorthWest Province*. Unpublished mimeograph.

Crowe, A. A. (2007a). *External moderation of Biology Senior Certificate CASS – Mpumalanga Province*. Unpublished mimeograph.

Crowe, A. A. (2009). *Composite report on DoE and IEB NSC Life Sciences question paper moderation 2009*. Unpublished mimeograph.

Crowe, A. A. (2010). Composite report on DBE and IEB NSC Life Sciences question paper moderation 2010. Unpublished mimeograph.

Crowe, A[lison]., Dirks. C., & Wenderoth, M. P. (2008). Biology in bloom: Implementing Bloom's Taxonomy to enhance student learning in Biology. *CBE Life Sciences Education, 7*(4), 368-381.

Cummins, J. (2003). BICS and CALP: Origins and rationale for the distinction. In C. B. Paulston & G. R. Tucker (Eds.), *Sociolinguistics: The essential readings* (pp. 322-328). London: Blackwell.

Dada, F., Dipholo, T., Hoadley, U., Khembo, E., Muller, S., & Volmink, J. (2009). *Report of the task team for the review of the implementation of the National Curriculum Statement*. Pretoria, South Africa: Department of Education.

Dall'Alba, G., & Edwards, J. (1983). Dimensions of cognitive demand. *Research in Science Education, 13*(1), 213-221.

Darling-Hammond, L. (1997). *The right to learn – A blueprint for creating schools that work.* San Francisco: Josse-Bass.

Darling-Hammond, L. (2003, February 16). Standards and assessments: Where we are now and what we need. *Teachers College Record,* ID No. 11109. Retrieved February 23, 2010, from http:/www.tcrecord.org

Darling-Hammond, L. (2007). Third Annual Brown Lecture in Education Research [−]The flat earth and education: How America's commitment to equity will determine our future. *Educational Researcher, 36*(6), 318-334.

Darling-Hammond, L., & Pecheone, R. (2007). *Developing an internationally comparable balanced assessment system that supports high-quality learning*. National conference on next generation assessment systems. Princeton, NJ: Educational Testing Service.

Department of Education. (1995). White paper on education and training. (WPJ-1995). *Government Gazette.*16312. Pretoria, South Africa: Government Printer.

Department of Education. (1996). *The National Education Policy Act 1996 (No. 27 of 1996*). Pretoria, South Africa: Author.

Department of Education. (1997a). *Curriculum 2005: Learning for the 21st century.* Pretoria, South Africa: Author.

Department of Education. (1997b*). A résumé of instructional programmes in public schools*. Report 550 (97/06). Pretoria, South Africa: Author.

Department of Education. (1998). *Investigation into the Senior Certificate Examination by the Ministerial Committee.* Pretoria, South Africa: Author.

Department of Education. (1999). *Call to action: Tirisano.* Pretoria, South Africa: Author.

Department of Education. (2000a). *Curriculum 2005: Towards a theoretical framework*. Pretoria, South Africa: Author.

Department of Education. (2000b). *Implementation plan for Tirisano, January 2000 – December 2004.* Pretoria, South Africa: Author.

Department of Education. (2000c). *Report on the international benchmarking of the 1999 Senior Certificate Examination.* Pretoria, South Africa: Author.

Department of Education. (2001a). *National Senior Certificate Examinations Guideline Document Biology H*[igher]*G*[rade] *& S*[tandard]*G*[rade] *Papers 1 & 2.* Pretoria, South Africa: Author.

Department of Education. (2001b). *General and Further Education and Training Quality Assurance Act No 58 of 2001.* Pretoria, South Africa: Author.

Department of Education. (2001c*). Principles related to marking H[igher]G[rade] & S[tandard]G[rade] Biology.* Quality improvement of the examinations and assessment system: Biology meeting. Pretoria, South Africa: Author.

Department of Education. (2001d*). A résumé of instructional programmes in schools.* Report 550 (2001/08*).* Pretoria, South Africa: Author.

Department of Education. (2001e). *[National] Senior Certificate Examination November 2001 Biology Higher Grade Paper 1.* Pretoria, South Africa: Author.

Department of Education. (2001f). *[National] Senior Certificate Examination November 2001 Biology Higher Grade Paper 2.* Pretoria, South Africa: Author.

Department of Education. (2001g). *[National] Senior Certificate Examination November 2001 Biology Standard Grade Paper 1*. Pretoria, South Africa: Author.

Department of Education. (2001h). *[National] Senior Certificate Examination November 2001 Biology Standard Grade Paper 2.* Pretoria, South Africa: Author.

Department of Education. (2002a). *National Senior Certificate Examinations Guideline Document Biology H*[igher]*G*[rade] *& S*[tandard]*G*[rade] *Papers 1 & 2: From 2002.* Pretoria, South Africa: Author.

Department of Education. (2002b). *Principles related to marking H[igher]G[rade] & S[tandard]G[rade] Biology. Quality improvement of the examinations and assessment system: Biology meeting.* Pretoria, South Africa: Author.

Department of Education. (2002c). *South Africa's National Research and Development Strategy.* Pretoria, South Africa: Author.

Department of Education. (2002d). *Education statistics in South Africa at a glance 2000*. Pretoria, South Africa: Author.

Department of Education. (2002e). *National Senior Certificate Examination November 2002 Biology Higher Grade Paper 1.* Pretoria, South Africa: Author.

Department of Education. (2002f). *National Senior Certificate Examination November 2002 Biology Higher Grade Paper 2.* Pretoria, South Africa: Author.

Department of Education. (2002g). *National Senior Certificate Examination November 2002 Biology Standard Grade Paper 1.* Pretoria, South Africa: Author.

Department of Education. (2002h). *National Senior Certificate Examination November 2002 Biology Standard Grade Paper 2.* Pretoria, South Africa: Author.

Department of Education. (2003a). *Principles related to marking H[igher]G[rade] & S[tandard]G[rade] Biology. Quality improvement of the examinations and assessment system: Biology meeting.* Pretoria, South Africa: Author.

Department of Education. (2003b). *Education statistics in South Africa at a glance 2001*. Pretoria, South Africa: Author.

Department of Education. (2003c). *National Senior Certificate Examination November 2003 Biology H[igher] G[rade] Paper 1.* Pretoria, South Africa: Author.

Department of Education. (2003d). *National Senior Certificate Examination November 2003 Biology H[igher] G[rade] Paper 2.* Pretoria, South Africa: Author.

Department of Education. (2003e). *National Senior Certificate Examination November 2003 Biology S[tandard] G[rade] Paper 1.* Pretoria, South Africa: Author.

Department of Education. (2003f). *National Senior Certificate Examination November 2003 Biology S[tandard] G[rade] Paper 2.* Pretoria, South Africa: Author.

Department of Education. (2004a). *National policy on the conduct, administration and management of the assessment for the Senior Certificate.* Pretoria, South Africa: Author.

Department of Education. (2004b). *Setting the national Biology examination papers: Style document*. Quality improvement of the examinations and assessment system: Biology meeting. Pretoria, South Africa: Author.

Department of Education. (2004c). *Rubrics for graphs, tables and drawings.* Quality improvement of the examinations and assessment system – Biology meeting. Pretoria, South Africa: Author.

Department of Education. (2004d). *Principles related to marking H*[igher]*G*[rade] *& S*[tandard]*G*[rade] *Biology*. Quality improvement of the examinations and assessment system: Biology meeting. Pretoria, South Africa: Author.

Department of Education. (2004e). *National strategy for mathematics, science and technology education. Pretoria, South Africa: Author.*

Department of Education. (2004f). *Education statistics in South Africa at a glance 2002*. Pretoria, South Africa: Author.

Department of Education. (2004g). *[National] Senior Certificate Examination October/November 2004 Biology Higher Grade P[aper]1.* Pretoria, South Africa: Author.

Department of Education. (2004h). *[National] Senior Certificate Examination October/November 2004 Biology Higher Grade P[aper]2.* Pretoria, South Africa: Author.

Department of Education. (2004i). *[National] Senior Certificate Examination October/November 2004 Biology Standard Grade P[aper]1.* Pretoria, South Africa: Author.

Department of Education. (2004j). *[National] Senior Certificate Examination October/ November 2004 Biology Standard Grade P[aper]2*. Pretoria, South Africa: Author.

Department of Education. [200[5]a]. *Senior Certificate Examination Report 2004*. Pretoria, South Africa: Author.

Department of Education. (2005b*). A résumé of instructional programmes in schools*. Report 550 (2005/09). *Government Gazette.*28159. Pretoria, South Africa: Government Printer.

Department of Education. (2005c). *Principles related to marking H[igher]G[rade] & S[tandard]G[rade] Biology. Quality improvement of the examinations and assessment system: Biology meeting.* Pretoria, South Africa: Department of Education.

Department of Education. (2005d). *Education statistics in South Africa at a glance 2003*. Pretoria, South Africa: Author.

Department of Education. (2005e*). [National] Senior Certificate Examination November 2005 Biology Higher Grade Paper 1.* Pretoria, South Africa: Author.

Department of Education. (2005f). *[National] Senior Certificate Examination November 2005 Biology Higher Grade Paper 2.* Pretoria, South Africa: Author.

Department of Education. (2005g). [National] Senior Certificate Examination November 2005 Biology Standard Grade Paper 1. Pretoria, South Africa: Author.

Department of Education. (2005h). [National] Senior Certificate Examination November 2005 Biology Standard Grade Paper 2. Pretoria, South Africa: Author.

Department of Education. (2006a). *Principles related to marking H[igher]G[rade] & S[tandard]G[rade] Biology.* Quality improvement of the examinations and assessment system: Biology meeting. Pretoria, South Africa: Department of Education.

Department of Education. (2006b). *Senior Certificate national examination results 2006*. Pretoria, South Africa: Author.

Department of Education. (2006c*). [National] Senior Certificate Examination October/November 2006 Biology Higher Grade P[aper]1*. Pretoria, South Africa: Author.

Department of Education. (2006d). *[National] Senior Certificate Examination October/November 2006 Biology Higher Grade P[aper]2.* Pretoria, South Africa: Author.

Department of Education. (2006e). *[National] Senior Certificate Examination October/November 2006 Biology Standard Grade P[aper]1).* Pretoria, South Africa: Author.

Department of Education. (2006f). *[National] Senior Certificate Examination October/November 2006 Biology Standard Grade P[aper]2.* Pretoria, South Africa: Author.

Department of Education. (2007a). *[National] Senior Certificate Examination October/ November 2007 Biology Higher Grade P[aper]1.* Pretoria, South Africa: Author.

Department of Education. (2007b). *[National] Senior Certificate Examination October/ November 2007 Biology Higher Grade P[aper]2.* Pretoria, South Africa: Author.

Department of Education. (2007c). *[National] Senior Certificate Examination October/ November 2007 Biology Standard Grade P[aper]1.* Pretoria, South Africa: Author.

Department of Education. (2007d). *[National] Senior Certificate Examination October/ November 2007 Biology Standard Grade P[aper]2.* Pretoria, South Africa: Author.

Department of Education. (2007e). *National Curriculum Statement Grades 10-12 (General) Subject Assessment Guidelines Life Sciences.* Pretoria, South Africa: Author.

Department of Education. (2009a). *Education Statistics in South Africa 2007*. Pretoria, South Africa: Author.

Department of Education. (2009b). *Life Sciences Examination Guidelines Grade 12 2009*. Pretoria, South Africa: Author.

Department of National Education. (1984a). *Core syllabus: Biology Higher Grade Standards 8* [Grade10]*, 9* [Grade 11] *and 10* [Grade 12]. Pretoria, South Africa: Author.

Department of National Education. (1984b). *Core syllabus: Biology Standard Grade Standards 8* [Grade10]*, 9* [Grade 11] *and 10* [Grade 12]. Pretoria, South Africa: Author.

Department of National Education. (1989*). A résumé of instructional programmes in public ordinary schools Nated 02- 550 (89/03).* Pretoria, South Africa: Author.

DeRoche, T. (2005). Beyond good and evil. *Teacher Magazine, 16*(5): 45-46. Retrieved February 27, 2005, from http://www.edweek.org

Dewey, J. (1938). *Experience and education.* New York: Simon & Schuster.

Dickson, M. (2005). *Further Education and Training (FET) and the National Curriculum Statement (NCS) Grades 10 –12.* Johannesburg, South Africa: Independent Schools Association of South Africa.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher, 35*(3), 125-129.

Domin, D. S. (1999). A content analysis of general chemistry laboratory manuals for evidence of higher-order cognitive tasks. *Journal of Chemical Education, 76*(1), 109-112.

Dorn, S. (1998). The policy legacy of school accountability systems. *Education Policy Analysis Archives, 6*(1). Retrieved February 17, 2006, from http://epa.asc/epaa/v6n13.html

Dowd, R. (2003). How national competency exams affect international political culture. *Journal of Public and International Affairs*, 14, 1 – 23.

Downing, S. T. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 3-25). Mahwah, NJ: Lawrence Erlbaum Associates.

Downing, S. T., & Haladyna, T. M. (1997). Test item development: Validity evidence from quality assurance procedures. *Applied Measurement in Education*, *10*(1)*,* 61-82.

Downing, S. T., & Haladyna, T. M. (Eds.) (2006). *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.

Drake, K. (1996). Comparison without comparability. In B. Boyle & T. Christie (Eds.), *Issues in setting standards: establishing comparabilities* (pp.42-56). London: Falmer Press.

Ebel, R. L. (1972). *Essentials of educational measurement.* New Jersey: Prentice Hall.

Eckstein, M. A., & Noah, H. J. (1993). *Secondary school examinations: International perspectives on policies and practice.* New Jersey, NJ: Yale University.

Edwards, J., & Dall'Alba, G. (1981). Development of a scale of cognitive demand for analysis of printed secondary science materials. *Research in Science Education, 11*(1)*,*158-170.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93*, 179-197.

Engelhard, G. (2005). *Measurement, trust and educational assessment.* Retrieved November 3, 2005, from http://www.bearcenter.berkley.edu/measurement/docs/CommentaryPorterEngelhard.pdf

Enger, S. K., & Yager, R. E. (2001). *Assessing student understanding in science: A standards-based K-12 handbook.* Thousand Oaks, CA: Corwin Press.

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity, 5*(3), 178-186.

Eubanks, D., & Eubanks, L. (1996). Chemistry examinations. In E. D. Britton & S. A. Raizen (Eds.), *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students* (pp. 87-120). Boston: Kluwer Academic Publishers.

Everitt, B. S., & Dunn, G. (1991). *Applied multivariate data analysis*. London: Edward Arnold.

Fadel, C., Honey, M., & Pasnik, S. (2007). Assessment in the age of innovation. *Education Week, 26*(38), 34;40. Retrieved May 23, 2007, from http://www.edweek.org/ew/articles/2007/05/23/38fadel.h26.html

Fatti, L. P. (2006). The statistical adjustment of marks. In V. Reddy (Ed.), *Marking Matric: Colloquium Proceedings* (pp. 45-57). Pretoria, South Africa: Human Sciences Research Council Press.

Feuer, M. J. (2010). Externalities of testing: Lessons from the blizzard of 2010. *Measurement: Interdisciplinary Research and Perspective*, *8*(2), 59-69.

Finn, C. E., Jr. (1990). The biggest reform of all. *Phi Delta Kappan*, 71(8), 584-592.

Ferrini-Mundy, J. (2004). *What does it mean to be standards-based? Issues in conceptualizing, measuring and studying alignment with standards.* SIRG Brief 3. Reston, VA: National Council of Teachers of Mathematics. Retrieved August 16, 2011, from http:// www.nctm.org/highered/sirg/sirg3.pdf

Fiske, E.B., & Ladd, H.F. (2006) 'Racial equity in education: How far has South Africa come?' *Perspectives in Education,* 24(2): 95-108.

Fitzpatrick, A. R., Ercikan, K.,Yen., W. M., & Ferrara, S. (1998). The consistency between raters scoring in different test years. *Applied Measurement in Education, 11*(2), 195-208.

Fleisch, B. (2008). Social order perspective on raising student achievement. *Journal of Educational Change. 9*(1), 53–70.

Fonthal, G. (2004). *Alignment of state assessments and higher education expectations: Definition and utilization of an alignment index.* Doctoral dissertation, University of California, Irvine and Los Angeles, CA.

Fowler, F. G., & Fowler, H. W. (Eds.) (1965). *The pocket Oxford dictionary of current English*. London: Clarendon Press.

Fraser, W. J., & Killen, R. (2003). Factors influencing academic success or failure of first year and senior university students: Do education students and lecturers perceive things differently? *South African Journal of Education, 23*(4), 254-260.

Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in Introductory Biology. *CBE Life Sciences Education, 10*(2), 175-186.

French, S. (1985). The weighting of examination components. *The Statistician, 34*(3), 265-280.

Frisbie, D. A. (2005). Measurement 101: some fundamentals revisited. *Educational Measurement, Issues and Practice, 24*(3), 21-28.

Fulmer, G. W. (2010). *Estimating critical values for strength of alignment among curriculum, assessments, and instruction.*. Paper presented at the 2010 Annual Meeting of the American Educational Research Association, Denver, United States of America. Retrieved October 13, 2010, from http://bensakovich.academia.edu/GavinFulmer/Papers/170595/_AERA_2010

Furst, E. J. (1981). Bloom's Taxonomy of Educational Objectives for the cognitive domain: Philosophical and educational issues. *Review of Educational Research, 51*(4), 441-453.

Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis, 19*(4), 325-338.

Gandal, M. (1994). What college-bound students abroad are expected to know about biology: A special report. *American Educator, 18*(1), 7-31.

Gandal, M., & McGiffert, L. (2003). The power of testing. *Educational Leadership, 60*(5), 39-42.

Gardner, H. (1998). Melding progressive and traditional perspectives. In M. S. Wiske (Ed.), *Teaching for understanding: Linking research with practice* (pp. 345-373). San Francisco: Jossey-Bass Publishers.

Gelman, A. (2003). Mechanistic understanding of models for educational assessment. *Measurement: Interdisciplinary Research and Perspective, 1*(1), 73-76.

Genovese, J. E. C. (2002). Cognitive skills valued by educators: Historical content analysis of testing in Ohio. *The Journal of Educational Research, 96*(2), 101-114.

Geisinger, K. F., & McCormick, C. M. (2010). Adopting cut scores: Post-standard-setting panel considerations for decision makers. *Educational Measurement: Issues and Practice, 29*(1), 38-44.

Gierl, M. J. (1997). Comparing cognitive representations of test developers and students on a mathematics test with Bloom's taxonomy. *The Journal of Educational Research, 91*(1), 26-32.

Gilmour, D., & Soudien, C. (2009). Learning and equitable access in the Western Cape, South Africa. *Comparative Education, 45*(2), 281-295.

Glas, C. A. W. (2003). On the structure of educational assessments: A comment from a psychometric perspective. *Measurement: Interdisciplinary Research and Perspective, 1*(1), 76-79.

Goldstein, H., & Cresswell, M. J. (1996). The comparability of different subjects in public examinations: A theoretical and practical critique. *Oxford Review of Education*, *22*(4), 435-442.

Gorin, J. S. (2007). Reconsidering issues in validity theory, *Educational Researcher, 36*(8), 456-462.

Govender, P. (2009, September 13). Plan to dumb down matric. *Sunday Times*, South Africa.

Government Communication & Information System, 2005/2006. *South Africa Yearbook 2005/2006 - Education.* Houghton, South Africa: STE Publishers. Retrieved March 12, 2007, from http://www.gcis.gov.za/docs/publications/yearbook/education.pdf

Granello, D. H. (2001). Promoting cognitive complexity in graduate written work: Using Bloom's Taxonomy as a pedagogical tool to improve literature reviews. *Counselor Education and Supervision, 40*(4), 292-307.

Greaney, V., & Kellaghan, T. (2008). *Assessing national achievement levels in education.* Washington, DC: The International Bank for Reconstruction and Development/The World Bank. Retrieved June 1, 2011, from http://siteresources.worldbank.org/EDUCATION/Resources/2782001099079877269/547664-1099079993288/assessing_national_achievement_level_Edu.pdf

Greatorex, J. (2003). *What happened to limen referencing? An exploration of how the awarding of public examinations has been and might be conceptualised.* Paper presented at the 2003 British Educational Research Association Conference, Edinburgh, Scotland. Retrieved November 3, 2010, from www.ucles-red.cam.ac.uk

Griesel, H. (2001). Institutional views and practices: Access, admissions, selection and placement. In *South African Universities Vice-Chancellors' Association (SAUVCA) – The challenges of access and admissions.* Pretoria, South Africa: SAUVCA.

Griffith, S. A. (2008). A proposed model for assessing quality of education. *International Review of Education, 54*, 99-112.

Griffiths, H. B., & Howson, A. G. (1974). *Mathematics, society and curricula.* Cambridge, England: Cambridge University Press.

Gross, P. R. (2009). Learning Science: Content – with reason. *American Educator, 33*(3), 35-40.

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in Introductory Biology. *Science, 332*, 1213-1216.

Haertel, E. H. (1999). Validity arguments for high stakes testing: In search of the evidence. *Educational Measurement: Issues and Practice,18*(4), 5-9.

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65-110). Westport, CT: Praeger Publishers.

Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking.* Needham Heights, MA: Allyn & Bacon.

Haladyna, T. M. (2006). Roles and importance of validity studies in test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 739-755). Mahwah, NJ: Lawrence Erlbaum Associates..

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and practice, 23*(1), 17-27.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement 15*(3), 309-334.

Hambleton, R. K. (1984). Validating the test score. In R. A. Berk (Ed.), *A guide to criterion-referenced test construction.* Baltimore, MD: John Hopkins University Press.

Hancock, G. R. (1994). Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. *Journal of Experimental Education, 62*(2), 143-157.

Hansche, L. N. (1998). *Handbook for the development of performance standards: Meeting the requirements of Title 1.* Washington, DC: Council of Chief State School Officers.

Harman, G. (1994). Student selection and admission to higher education: Policies and practice in the Asia region. *Higher Education*, *27*(3), 313 -339.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal, 113,* 64-98.

Häussler, P., & Hoffmann, L. (2000). A curricular frame for physics education: Development, comparison with students' interests, and impact on students' achievement and self-concept. *Science Education*, *84*(6), 689-705.

Herman, H. D. (1995). School-leaving examinations, selection and equity in higher education in South Africa. *Comparative Education, 31*(2): 261-274.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Herman, J. L., & Webb, N. M. (2007). Guest editors introduction: Alignment methodologies. *Applied Measurement in Education, 20*(1), 1-5.

Herr, N. (1992). A comparative analysis of the perceived influence of Advanced Placement and Honors Programs upon science instruction. *Journal of Research in Science Teaching, 29*(5), 521-532.

Hess, F. M. (2010). *The same things over and over: How school reformers get stuck in yesterday's ideas.* Cambridge, MA: Harvard University Press.

Heyneman, S. P. (2004). International education quality. *Economics of Education Review, 23*(4), 441-452.

Hiebert, J. (1999). Relationships between research and the NCTM standards. *Journal for Research in Mathematics Education, 30*(1), 3-19.

Hill, H. C. (2001). Policy is not enough: Language and the interpretation of state standards. *American Educational Research Journal, 38*(2), 289-318.

Hill, P. W., & McGaw, B. (1981).Testing the simplex assumption underlying Bloom's 'Taxonomy'. *American Educational Research Journal, 18*(1), 93-101.

Hindle, D. (2009, August 18). We must not condemn our children. *Cape Times,* South Africa.

Hirsch, E. D. (1996). *The schools we need and why we don't have them*. New York: Anchor Books, Random House.

Hirsch, E. D. (2001).Seeking breadth and depth in the curriculum. *Educational Leadership 59*(2), 22-25.

Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.187-220). Westport, CT: Praeger Publishers.

House of Representatives Department of Education (1986). *Senior Secondary Course Syllabus Biology (Higher Grade) Standards 8, 9 and 10*[Grades 10, 11 and 12]- *SS 9/86.*Cape Town, South Africa: Author.

Huff, K., & Plake, B. S. (2010). Innovations in setting performance standards for K-12 test-based accountability. *Measurement: Interdisciplinary Research and Perspective*, *8*(2), 130-144.

Humphry, S. M. (2011): The Role of the Unit in Physics and Psychometrics. *Measurement: Interdisciplinary Research and Perspective, 9*(1), 1-24.

Impara, J. C. (2001). *Alignment: One element of an assessment's instructional unity.* Paper presented at the 2001 Annual Meeting of the National Council on Measurement Education, Seattle, WA. Retrieved June 17, 2005, from http//www.unl.edu/BIACO/NCME/Alignment%20revised.pdf

Impara, J. C., & Plake, B. S. (1998). Teachers' ability to estimate item difficulty: A test of the assumptions of the Angoff standard setting method. *Journal of Educational Measurement*, *35*(1), 69-81.

Independent Examinations Board (1996). *IEB Resource material for Biology 2 − A guide to the syllabus: Std 10.[Grade 12].* Johannesburg, South Africa: Author.

Independent Examinations Board (2001a). *IEB Senior Certificate Examination November 2001 Biology Higher Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board(2001b). *IEB Senior Certificate Examination November 2001 Biology Standard Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2002a). *IEB Senior Certificate Examination November 2002 Biology Higher Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2002b). *IEB Senior Certificate Examination November 2002 Biology Standard Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2003a). *IEB Senior Certificate Examination November 2003 Biology Higher Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2003b). *IEB Senior Certificate Examination November 2003 Biology Standard Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2004a*). IEB Senior Certificate Examination November 2004 Biology Higher Grade*. Johannesburg, South Africa: Author.

Independent ExaminationsBoard (2004b). *IEB Senior Certificate Examination November 2004 Biology Standard Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2005a). *IEB Senior Certificate Examination November 2005 Biology Higher Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2005b). *IEB Senior Certificate Examination November 2005 Biology Standard Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2006a). *IEB Senior Certificate Examination November 2006 Biology Higher Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2006b). *IEB Senior Certificate Examination November 2006 Biology Standard Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2007a). *IEB Subject Assessment Guidelines − Life Sciences 2008.* Johannesburg, South Africa: Author.

Independent Examinations Board (2007b). *IEB Senior Certificate Examination November 2007 Biology Higher Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2007c). *IEB Senior Certificate Examination November 2007 Biology Standard Grade.* Johannesburg, South Africa: Author.

Independent Examinations Board (2009). *Life Sciences: IEB National Senior Certificate Handbook – Grade 12 January 2010* (pp. 24/1- 24/32). Johannesburg, South Africa: Author.

Indge, B. (1997). *The complete A−Z Biology handbook.* London: Hodder & Stroughton.

Isaac, T. (1990). *The influence of the Senior Certificate Examination on methods of teaching Biology at Standard 10* [Grade 12] *level in Indian secondary schools.* Doctoral dissertation, University of South Africa, Pretoria, South Africa.

Ivie, S. D. (1998). Ausubel's learning theory: An approach to teaching higher order thinking skills. *The High School Journal, 82*(1), 35-42.

Jackson, I. M., & Young, D. A. (1987). Trends in the relationship between matriculation results and success in first-year biology studies at university. *South African Journal of Education, 7*(2), 132-136.

Jaeger, R. M. (1991). Selection of judges for standard setting. *Educational Measurement, Issues and Practice, 10*(2), 3-6, 10, 14.

Janofsky, M. (2005, December 8). Report says state standards aim low in science classes. *New York Times.* Retrieved December 14, 2005, from http://www.nytimes.com/2005/12/08/education/08schools.html

James, F. C., & McCulloch, C. E. (1990). Mutivariate analysis in ecology and systematics: Panacea or Pandora's box? *Annual Review of Ecology and Systematics, 21*, 129-166.

Jansen, J. D. (1999). The school curriculum since apartheid: Intersections of politics and policy in the South African transition. *Journal of Curriculum Studies, 31*(1), 57-67.

Jansen, J. D. (2001). Explaining non-change in education reform after apartheid: political symbolism and the problem of policy implementation. In J. D. Jansen & Y. Sayed (Eds.), *Implementing education policies: the South African experience* (pp. 271- 292). Cape Town, South Africa: Juta & Company Ltd.

Jansen, J. [D.] (2012). National Senior Certificate results belie conceptual and skill limitations of school-leavers. *South African Journal of Science, 108*(3/4), 1-2.

Joint Matriculation Board (1989). *Matriculation examination handbook 1989 and 1999 rules and regulations*[,] *prescribed works*[,] *acts and statutes.* Pretoria, South Africa: Author.

Jones, B. E. (1997). Comparing examination standards: Is a purely statistical approach adequate? *Assessment in Education*: *Principles, Policy & Practice, 4*(2), 249-263.

Jones, B., & Ratcliffe, P. (1996). Comparing the standards of examining groups in the United Kingdom. In B. Boyle & T. Christie (Eds.), *Issues in setting standards: Establishing comparabilities* (pp. 57-70). London: Falmer Press.

Jones, P., Smith, R. W., & Talley, D. (2006). Developing test forms for small-scale achievement testing systems. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 487-525). Mahwah, NJ: Lawrence Erlbaum Associates..

Kafatos, F. C., & Eisner, T. (2004). Unification in the century of biology. *Science, 303*, 1257.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527-535.

Kane, M. T. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research, 64*(3), 425-461.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*(4), 319-342.

Kane, M. T. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice, 21*(1), 31-41.

Kane, M. T. (2006a). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.17-64). Westport, CT: Praeger Publishers.

Kane, M. T. (2006b). Content-related validity evidence in test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 131-153).Mahwah, NJ: Lawrence Erlbaum Associates..

Kane, M. [T.] (2009). Validating the interpretations and uses of test scores. In R.W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 39-64). Charlotte, NC: Information Age Publishing.

Kanjee, A. (2006). Comparing and standardising performance trends in the matric examinations using matrix sampling design. In V. Reddy (Ed.), *Marking Matric: Colloquium Proceedings* (pp. 72-89). Pretoria, South Africa: Human Sciences Research Council Press.

Karamustafaoğlu, S., Sevįm, S., Karamustafaoğlu, O., & Çepnį, S. (2003). Analysis of Turkish high-school Chemistry-examination questions according to Bloom's Taxonomy. *Chemistry Education: Research and Practice, 49*(1), 25-30.

Keeves, J. P. (1994). *National examinations: Design, procedures and reporting*. Paris: United Nations Educational, Scientific and Cultural Organization. Retrieved June 14, 2010, from http://unesdoc.unesco.org/images/0009/000992/099209eo.pdf

Kellaghan, T. (1992). Examination systems in Africa: Between internalization and indigenization. In M. A. Eckstein & H. J. Noah (Eds.), *Examinations: comparative and international studies* (pp. 95-104). Oxford, England: Pergamon.

Kendall, J. S., Ryan, S. E., & Richardson, A. T. (2005). *The systematic identification of performance standards*. (US Department of Education contract number ED-01-CO-0006). Denver, CO: McREL. Retrieved July 6, 2010, from https://www.mcrel.org/PDF/standards/5041TG_PerfStnds.pdf

Killen, R. (2003). Validity in outcomes-based assessment. *Perspectives in Education, 21*(1), 1-14.

Kim, S., Walker, M. E., & McHale, F. (2009). Equating mixed-format tests in large-scale assessments. *ETS Research Spotlight, 2*, 15-20.

Klein, M. F. (1972). Use of the Taxonomy of Educational Objectives (Cognitive Domain) in constructing tests for primary school pupils. *Journal of Experimental Education*, *40*(3), 38-50.

Kliebard, H. M. (1970). The Tyler Rationale. *The School Review, 78*(2), 259-272.

Knapp, C. E. (1992). *Thinking in outdoor inquiry.* Charleston, West Virginia: ERIC Clearinghouse on Rural Education and Small Schools (Eric Document Reproduction Service No. ED 348 198). Retrieved June 15, 2007, from http://www.ed.gov/databases/ERIC_Digests/ed348198.html

Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices.* New York: Springer-Verlag.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating,, scaling and linking: Methods and practices.* New York: Springer-Verlag.

Kolen, M. J. (2006). Scaling and norming. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.155-186). Westport, CT: Praeger Publishers.

Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: An overview. *Theory Into Practice, 41*(4), 212-218.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook 2: Affective domain.* London: Longman Group.

Kreitzer, A. E., & Madaus, G. F. (1994). Empirical investigations of the hierarchical structure of the Taxonomy. In L. W. Anderson, & L. A. Sosiak (Eds.), *Bloom's Taxonomy: A forty-year retrospective, Ninety-third yearbook of the National Society for the Study of Education*, *Part II* (pp. 64-81). Chicago: University of Chicago Press.

Kropp, R. P., Stoker, H. W., & Bashaw, W. L. (1966). The validation of the Taxonomy of Educational Objectives. *The Journal of Experimental Education, 34*(3), 69-76.

Kunen, S., Cohen, R., & Solman, R. (1981). A levels-of-processing analysis of Bloom's Taxonomy. *Journal of Educational Psychology, 73*(2), 202-211.

KwaZulu-Natal Department of Education (1996). *Interim Core Syllabus and Provincialised Guide for Biology [Std] 10* (*Higher Grade) 1996*. Pietermaritzburg, South Africa: Author.

KwaZulu-Natal Department of Education (1998). *Examiner's report: Biology Higher Grade December 1997*. Unpublished mimeograph.

La Marca, P. M. (2001). Alignment of standards and assessments as an accountability criterion. *Practical Assessment, Research & Evaluation*, 7(21). Retrieved June 18, 2007, from http://pareonline.net/getvn.asp?v=7&n=21

La Marca, P. M., Redfield, D., Winter, P. C. Bailey, A., & Despriet, L. (2000). *State standards and state assessment systems: A guide to alignment.* Washington, DC: Council of Chief State School Officers.

Lane, S. (2004). NCME Presidential Address – Validity of high-stakes assessment: Are students engaged in complex thinking? *Educational Measurement: Issues and Practice, 23*(3), 6-14.

Lane, S. (2011). Practical issues in the assessment of higher order thinking skills. In G. Schraw & D. R. Robinson (Eds.), *Assessment of higher order thinking skills* (pp. 265-302). Charlotte, NC: Information Age Publishing.

Larson, E. D. (2003). *Aligning science assessment items from the Iowa Testing Program Batteries and the State Collaborative on Assessment and Student Standards with the National Science Education Content Standards.* Doctoral dissertation, University of Iowa, Iowa City, IA.

Laugksch, R. C. (2001). The differential role of Physical Science and Biology in achieving scientific literacy amongst selected matriculants – a possible explanation. *South African Journal of Education, 21*(3)*,* 157 – 161.

Lee, O. (2005). Science education with English language learners: Synthesis and research agenda. *Review of Educational Research, 75*(4), 491-530.

Leighton, J. P. (2011). A cognitive model for the assessment of higher order thinking in students. In G. Schraw & D. R. Robinson (Eds.), *Assessment of higher order thinking skills* (pp. 151-181). Charlotte, NC: Information Age Publishing.

Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement, 41*(5)*,* 205-236.

Letseka, M., & Maile, S. (2008, March). *High university drop-out rates: a threat to South Africa's future.* Human Sciences Research Council Policy Brief. Cape Town, South Africa: Human Sciences Research Council Press.

Lewin, K., & Dunne, M. (2000). Policy and practice in assessment in anglophone Africa: Does globalisation explain convergence? *Assessment in Education, 7*(3), 379-399.

Li, M. (2002). *A framework for science achievement and its link to test items*. Doctoral dissertation. University of Stanford, CA.

Linn, R. L. (1993). Educational Assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis, 15*(1), 1-16.

Linn, R. L. (1995). High-stakes uses of performance-based assessments: Rationale, examples, and problems of comparability. In T. Oakland, & R. K. Hambleton (Eds.), *International perspectives on academic assessment* (pp. 49-74). Norwell, MA: Kluwer Academic Publishers.

Linn, R. L. (2001). *Standards-based accountability – ten suggestions.* CRESST Policy Brief. Los Angeles: UCLA Graduate School of Education & Information Studies. Retrieved February 17, 2006, from http://www.cse.ucla.edu/products/newsletters/policypaper.pdf

Linn, R. L. (2003). Performance standards: Utility for different uses of assessments. *Education Policy Analysis Archives, 11*(31), 1-20.

Linn, R. L. (2006). The standards for educational and psychological testing: Guidance in test development. In S. M. Downing & T. M. Haladyna (Eds.) *Handbook of test development* (pp. 27-38). Mahwah, NJ: New Jersey: Lawrence Erlbaum Associates.

Linn, R. L. (2009). The concept of validity in the context of NCLB. In R.W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 195-212). Charlotte, NC: Information Age Publishing.

Linn, R. L., & Baker, E. L. (1995). What do international assessments imply for world-class standards? *Educational Evaluation and Policy Analysis, 17*(4), 405-418.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, *20*(8), 5-21.

Linn, R. L., & Herman, J. L. (1997). *A policy maker's guide to standards-led assessment.* Denver, CO: Education Commission of the States. Retrieved March 16, 2007, from http://www.ecs.org/clearinghouse/29/14/2914.doc

Lissitz, R.W., & Samuelsen, K. (2007a). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher, 36*(8), 437-448.

Lissitz, R.W., & Samuelsen, K. (2007b). Further clarification regarding validity and education. *Educational Researcher, 36*(8), 482-484.

Lissitz, R. W. (2009). Introduction. In R.W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 1-15). Charlotte, NC: Information Age Publishing.

Little, A. (1996). Contexts and histories: the shaping of assessment practice. In A. Little & A. Wolf (Eds.), *Assessment in transition: learning, monitoring and selection in international perspective* (pp. 3-27). Oxford, England: Elsevier.

Liu, X., & Fulmer, G. (2008). Alignment between the science curriculum and assessment in selected NY State Regents Exams. *Journal of Science Education and Technology*, 17(4), 373–383.

Liu, X., Zhang, B., Liang, L. L., Fulmer, G., Kim, B., & Yuan, H. (2009). Alignment between the physics content standards and the standardized test: A comparison among the United States-New York State, Singapore, and China-Jiangsu. *Science Education*, 93(5), 777-797.

Livingston, S. A. (2004) *Equating test scores (without IRT)*. Princeton, NJ: Educational Testing Service.

Livingston, S. A. (2006). Item analysis. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 421-444). Mahwah, NJ: Lawrence Erlbaum Associates.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*(Monograph Supplement 9), 635-694.

Lolwana, P. (2006). The history of falling matric standards. In V. Reddy (Ed.), *Marking Matric: Colloquium Proceedings* (pp. 18-32). Pretoria, South Africa: Human Sciences Research Council Press.

Long, V. M. & Benson, C. (1998). Re: Alignment. *The Mathematics Teacher, 91*(6), 503-508.

Lord, T. R. (1997). A comparison between traditional and constructivist teaching in college biology. *Innovative Higher Education, 21*(3), 197-216.

Lord, T., & Baviskar, S. (2007). Moving students from information recitation to information understanding: Exploiting Bloom's Taxonomy in creating science questions. *Journal of College Science Teaching, 36*(5), 40-44.

Louis, K. S., & Versloot, B. (1996). High standards and cultural diversity: Cautionary tales of comparative research − A comment on "Benchmarking education standards" by Lauren B. Resnick, Katherine J. Nolan, and Daniel P. Resnick. *Educational Evaluation and Policy Analysis, 18*(3), 253-261.

Loveless, T. (2012). How well are American students learning. *Brown Center Report on American Education, 3*(1), 1-33. Retrieved February 19, 2012, from http://www.brookings.edu/~/media/Files/rc/reports/2012/0216_brown_education_loveless/0216_brown_education_loveless.pdf

Lubisi, R. C., & Murphy, R. J. L. (2002). Assessment in South African schools. A*ssessment in Education*, *9*(2), 255-268.

Luckay, M. B. (2010). Implementation of social constructivist learning environments in Grade 9 Natural Science in the Western Cape Province, South Africa. Doctoral dissertation, University of Cape Town, Cape Town, South Africa.

Luke, A. (2011). Generalizing across borders: Policy and the limits of educational science. *Educational Researcher, 40*(8), 367-377.

Madaus, G. F. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.), *Critical issues in curriculum: Eighty-seventh yearbook of the National Society for the Study of Education, Part 1*(pp. 83-121). Chicago, IL: University of Chicago Press.

Madaus, G. F., Woods, E. M., & Nuttall, R. L. (1973). A causal model analysis of Bloom's Taxonomy. *American Educational Research Journal, 10*(4), 253-262.

Maile, S. (2004). School choice in South Africa. *Education and Urban Society, 37*(1), 94-116.

Malan, S. P. T. (2000). The 'new paradigm' of outcomes-based education in perspective. *Journal of Family Ecology and Consumer Sciences, 2,* 22-28. Retrieved September 9, 2010, from http://ajol.info/index.php/jfecs/article/viewFile/52788/41390

Mansilla, V. B., & Gardner. H. (1998). What are the qualities of understanding? In M. S. Wiske (Ed.), *Teaching for understanding: Linking research with practice* (pp. 161-196). San Francisco: Jossey-Bass Publishers.

Marshall, J. C., & Horton, R. M. (2011). The relationship of teacher-failitated inquiry-based instruction tostudent higher-order thinking. *School Science and Mathematics, 111*(3), 93-101.

Martin, M., Mullis, I., Gonzalez, E. J., Gregory, K. D., Smith, T., Chrostowski, S. J., Garden, R. A., & O'Connor, K. M. 2000. *TIMSS 1999 International Science Report: Findings from IEA*'s *repeat of the Third International Mathematics and Science Study at eighth grade.* Chestnut Hill, MA: Boston College.

Martinez, M. E. (1999). Cognition and the question of item test format. *Educational Psychologist, 34*(4)*,* 207-218.

Martiniello, M. (2008). Language and the performance of English language learners in math word problems. *Harvard Educational Review, 78*(2), 333-368.

Martiniello, M. (2009). Lingusitic complexity, schematic representations, and differential item functioning for English language learners in maths tests. *Educational Assessment, 14*(3/4), 160-179.

Marzano, R. J. (2001). *Designing a new taxonomy of educational objectives.* Thousand Oaks, CA: Corwin Press.

Marzano, R. J., & Costa, A. L. (1988). Question: Do standardized tests measure general cognitive skills? Answer: No. *Educational Leadership, 45*(8), 66-7.

Marzano, R. J., & Kendall, J. S. (1997, September). *The fall and rise of standards-based education*. National Association of State Boards of Education (NASBE) Issues in Brief. Retrieved November 5, 2010, from http://fc.sad57.k12.me.us/~ci_a/5962IR_FallAndRise.pdf

Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice,* 42(4), 226-232.

McClure,P. (2005). Where standards come from. *Theory into Practice, 44*(1), 4-10.

McConnell, D. A., Steer, D. N., & Owens, K. D. (2003). Assessment and active learning strategies for introductory geology courses. *Journal of Geoscience Education, 51*(2), 205-216.

McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis, 17*(3), 305-322.

McGinty, D. (2005). Illuminating the "Black Box" of standard setting: An exploratory qualitative study. *Applied Measurement in Education, 18*(3), 269-287.

McMillan, J. H., & Schumacher, S. (2001). *Research in education – A conceptual introduction.* New York: Longman.

McNamara, T. (2003). Knowing your limits: Conceptual and policy constraints on evidence-sentred test design in language testing research. *Measurement: Interdisciplinary Research and Perspective, 1*(1), 85-87.

Meagher, T. R., & Futuyma, D. J. (2001). Forward: Evolution in the century of Biology. *American Naturalist, 158*(S4), 1-46.

Messick, S. (1989a). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp.13-103). Westport, CT: Praeger Publishers.

Messick, S. (1989b). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Metcalf, M., Orkin, M., & Glennie, J. (2012, February 15). Our pass rate focus is too narrow. *Sunday Times,* South Africa.

Meyers,J. L., Miller, G. E., & Way, W. D. (2009). Item position and item difficulty change an IRT-based common item equating design. *Applied Measurement in Education, 22*(1), 38-60.

Michell, J. (1999). *Measurement in psychology*. Cambridge, United Kingdom: Cambridge University Press.

Michell, J. (2008). Is psychometrics pathological science? *Measurement: Interdisciplinary Research and Perspective,6*, 7–24.

Miller, H. G., Williams, R. G., & Haladyna, T. M. (1978). *Beyond facts: Objective ways to measure thinking.* Englewood Cliffs, NJ: Educational Technology Publications.

Miri, B., David, B., & Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science Education, 37,* 353-369.

Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods and prospects.* Princeton, NJ: ETS Policy Information Centre.

Mislevy, R. J. (1995). What can we learn from international assessments? *Educational Evaluation and Policy Analysis 17*(4), 419-437.

Mislevy, R. J. (2008). *Some implications of expertise research for educational assessment*. Keynote address at the 34th International Association for Educational Assessment (IAEA) Conference, Cambridge University, September 8, 2008. Retrieved May 8, 2011, from http://www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/180456_Mislevy.pdf

Mislevy, R. J. (2009). Validity from the perspective of model-based reasoning. In R.W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 83-108). Charlotte, NC: Information Age Publishing.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspective, 1*(1), 3-62.

Mitchell, G., & Fridjhon, P. (1987). Matriculation examinations and university performance. *South African Journal of Science*, *83*(4), 555-559.

Mitchell, G., Fridjhon, P., & Haupt, J. (1997). On the relationship between the matriculation examination and university performance in South Africa 1980-1991. *South African Journal of Science*, *93*(9), 382-386.

Moreno, N. P., & Tharp, B. Z. (2006). How do students learn science?. In J. Rhoton & P. Shane (Eds.), *Teaching science in the 21st century* (pp. 291- 305). Arlington, Va: National Science Teachers Association Press.

Moseley, D., Baumfield, V., Elliott, J., Gregson, M., Higgins, S., Miller, J., & Newton, D. 2005. *Frameworks for thinking – A handbook for teaching and learning.* Cambridge, England: Cambridge University Press.

Moss, P. A. (2004). The risks of coherence. *Yearbook of the National Society for the Study of Education*, *103*(2), 217-238.

Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education, 30*(Special issue), 109-162.

Mukwevho, T., Khosa, V. R., & Kgobe, M. (2004). An analysis of trends in matric results. *Quarterly Review of Education and Training in South Africa* 11(1), 19-23.

Muller, J. (2004). Assessment, qualifications and the NQF in South African schooling. In L. Chisholm (Ed.), *Changing class – Education and social change in post-apartheid South Africa* (pp. 221-246). Cape Town, South Africa: HSRC Press.

Muller, J. (2005). The challenge of cognitive demand. In *Matric: What is to be done?* (pp. 40-44). Seminar papers and presentations – Umalusi and the Centre for Higher Education Transformation Seminar on Matric. Pretoria, South Africa: Umalusi.

Murphy, R. (2003). A review of the changing views of public examinations. In *Setting the standard* (pp. 181-189). Manchester, England: The Assessment and Qualifications Alliance.

Naidoo, M. (2006). Transition from Senior Certificate to the Further Education and Training Certificate. In V. Reddy (Ed.), *Marking Matric: Colloquium Proceedings* (pp.10-17). Pretoria, South Africa: Human Sciences Research Council Press.

Naidoo, P., & Lewin, K. M. (1998). Policy and planning of Physical Science education in South Africa: Myths and realities. *Journal of Research in Science Teaching, 35*(7), 729-744.

Naidu, E. (2006, 8 January). Prospects grim for the class of 2005. *The Sunday Independent*, South Africa.

National Academy of Sciences. 1998. *Teaching about evolution and the nature of science*. Washington, DC: National Academies Press.

National Education Policy Investigation. (1992a). *Curriculum.* Cape Town, South Africa: Oxford University Press/National Education Co-ordinating Committee.

National Education Policy Investigation. (1992b). *The framework report and final summaries.* Cape Town, South Africa: Oxford University Press/National Education Co-ordinating Committee

National Education Policy Investigation. (1992c). *Governance and administration.* Cape Town, South Africa: Oxford University Press/National Education Co-ordinating Committee.

National Governors Association Center for Best Practices, & Council of Chief State School Officer Common Core State Standards Initiative. (2010). *Standards-setting criteria.* Washington DC: Author. Retrieved March 2, 2010, from www.corestandards.org/Files/Criteria.pdf

National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academies Press.

National Research Council. (1997). *Improving student learning in mathematics and science: The role of national standards in state policy.* Washington DC: National Academies Press.

National Research Council. (1999). *How people learn: Brain, mind, experience and school.* Washington DC: National Academies Press.

National Research Council. (2002). *Investigating the influence of standards.* Washington, DC: National Academies Press.

National Research Council. (2006). *Systems for state science assessment.* Committee on Test Design for K-12 Science Achievement. M. R. Wilson, & M.W. Bertenthal, (Eds.). Washington, DC: National Academies Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington DC: National Academies Press.

National Research Council. (2008). *Common standards for K-12 education? Considering the evidence: Summary of a workshop series.* Washington, DC: National Academies Press.

National Research Council. (2012). *A framework for K-12 Science Education: Practices, crosscutting concepts, and core ideas.* Washington, DC: National Academies Press.

National Science Teachers Association. (1982). *An NSTA position statement*. Washington, DC: Author.

Ndlovu, T., Sishi, N. & Deliwe, C. N. (2006). A review of ten years of assessment and examinations. In V. Reddy (Ed.), *Marking Matric: Colloquium Proceedings* (pp.3-9). Cape Town: Human Sciences Research Council Press.

News24.com, (2003, December 31). *2003 matric pass rate 73.3%.* Retrieved February 13, 2005, from http://www.southafrica.info/ess_info/sa_glance/education/matricresults03.htm

Newmann, F. M., Smith, B., Allensworth, E., Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis, 23*(4), 297-322.

Newton, P. (1997). Examining standards over time. *Research Papers in Education, 12*(3), 227-248.

Newton, P. (2007). *Techniques for monitoring the comparability of examination standards.* Paper presented at the International Association for Educational Assessment 33rd Annual Conference, Baku, Azerbaijan. Retrieved November 3, 2010, from http://iaea2007.tqdk.gov.az/cp/Techniques%20for%20monitoring%the%20comparability%20of%20examination%20standards.pdf

Nichols, P., Twing, J., Mueller, C. D., & O'Malley, K. (2010). Standard-setting methods as measurement processes. *Educational Measurement: Issues and Practice, 29*(1), 14-24.

Nitko, A. J. (2001). *Educational assessments of students* (3rd ed.). Upper Saddle, NJ: Prentice-Hall/Merrill.

Nitko, A. J., & Brookhart, S. M. (2007). *Educational assessments of students* (5th ed.). Upper Saddle River, NJ: Prentice-Hall/Merrill.

Noah, H. J. (1989). An international perspective on national standards. *Teachers College Record. 91*(1): 17-22.

Noah, H. J. (1996). One best system? Lessons from comparative research for assessment policy. In A. Little & Wolf, A. (Eds.), *Assessment in transition: Learning, monitoring and selection in international perspective* (pp. 88-95). Oxford, England: Pergamon.

Nordvall, R. C., & Braxton, J. M., (1996). An alternative definition of undergraduate college education. Toward usable knowledge for improvement. *The Journal of Higher Education, 67*(5), 483-497.

Nuttall, D. L. (1979). The myth of comparability. *Journal of the National Association of Inspectors and Advisors,* 11, 16-18.

Nuttall, D. L. (1986). Problems in the measurement of change. In D. L. Nuttall, (Ed.), *Assessing Educational Achievement* (pp.118-131). London: Falmer Press.

Oakland, T. (1995). Test use with children and youth internationally: Current status and future directions. In T. Oakland, & R. K. Hambleton (Eds.), *International perspectives on academic assessment* (pp. 1- 24). Norwell, MA: Kluwer Academic Publishers.

Oberholzer, A. K. (1995). An investigation into the comparability of Senior Certificate examination results in different subjects and at different grades. Master's dissertation, University of Natal, Durban, South Africa.

O'Neil, T., Sireci, S. G., & Huff, K. L. (2003-2004). Evaluating the consistency of test content across two successive administrations of a state-mandated science assessment. *Educational Assessment, 9*(3&4), 129-151.

Organisation for Economic Co-operation and Development. (2008). *Reviews of National Policies for Education: South Africa.* Paris: Author.

Orkin, F. M. (2006). Foreword. In V. Reddy (Ed.), *Marking Matric: Colloquium Proceedings* (pp. xii-xix). Cape Town: Human Sciences Research Council Press.

Ormell, C. P. (1974). Bloom's Taxonomy and the objectives of education. *Educational Research, 17*(1), 3-18.

Ormell, C. P. (1979). Problems of analyzing understanding. *Educational Research, 22*(1), 32-38.

Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "ideas-about-science" should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching 40*(7), 692-720.

Paterson, A. (2003). Standards. In *Setting the standard (*pp 139-158). Manchester, England: The Assessment and Qualifications Alliance.

Pellegrino, J.W., Chudowsky, N., & Glaser, R. (Eds.) (2001). *Knowing what students know: The science and design of educational assessment.* Board on Testing and Assessment, Center for Education, National Research Council. Washington, DC: National Academies Press.

Penfield, R. D., & Lee, O., 2010. Test-based accountability: Potential benefits and pitfalls of science assessment with student diversity. *Journal of Research in Science Teaching, 47*(1), 6-24.

Penick, J. E., & Harris, R. L. (2005). *Teaching with purpose: Closing the research-practice gap*. Arlington, VA: National Science Teachers Association Press.

Perkins, D. (1993a). Creating a culture of thinking. *Educational Leadership, 51*(3), 98-99.

Perkins, D. (1993b). Teaching for understanding. *American Educator, 17*(3), 28-35.

Perkins, D. (1998a). What is understanding? In M. S. Wiske (Ed.), *Teaching for understanding: Linking research with practice* (pp. 39-57). San Francisco: Jossey-Bass Publishers.

Perkins, D. (1998b). Conclusion: Melding progressive and traditional perspectives. In M. S. Wiske (Ed.), *Teaching for understanding: Linking research with practice* (pp. 345-350). San Francisco: Jossey-Bass Publishers.

Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher, 18*(1), 16-25.

Perrone, V. (1998). Why do we need a pedagogy of understanding? In M. S. Wiske (Ed.), *Teaching for understanding: Linking research with practice* (pp. 13-38). San Francisco: Jossey-Bass Publishers.

Phurutse, M. C. (2006). Listening to matric teachers: Township realities and learn achievement levels. In V. Reddy (Ed.), *Marking Matric: Colloquium Proceedings* (pp. 213-227). Pretoria, South Africa: Human Sciences Research Council Press.

Pitoniak, M. J., Young, J. W., Martiniello, M., King, T. C., Buteux, A., & Ginsburgh, M. (2009). *Guidelines for the assessment of English-Language Learners.* Princeton, NJ: Educational Testing Service.

Poliah, R. R. (2009). The management of the quality assurance of school based assessment at a national level in South Africa. Doctoral dissertation, University of Johannesburg, Johannesburg, South Africa.

Popham, W. J. (2004). Curriculum, instruction, and assessment: amiable allies or phony friends? *Teachers College Record, 106*(3), 417-428.

Popham, W. J. (2006). Needed: A dose of assessment literacy. *Educational Leadership, 63*(6), 84-85.

Popham, W. J. (2009). *Unlearned lessons* [−] *Six stumbling blocks to our schools' success*. Cambridge, MA: Harvard Education Press.

Porter, A. C. (1989). External standards and good teaching: The pros and cons of telling teachers what to do. *Educational Evaluation and Policy Analysis, 11*(4), 343-356.

Porter, A. C. (1991). Creating a system of school process indicators. *Educational Evaluation and Policy Analysis, 13*(1), 13-29.

Porter, A. C. (1995). The uses and misuses of opportunity to learn standards. *Educational Researcher, 24*(1), 21-27.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, *31*(7), 3-14.

Porter, A. C. (2006). Curriculum Assessment. In J. L. Green, G. Camilli & P. B. Elmore (Eds.), *Handbook of* complementary methods in education research* (pp. 141-159). Washington, DC: American Educational Research Association. Retrieved November 24, 2005, from http://www.vanderbilt.edu/lsi/ expert/documents/vitae/porter.pdf

Porter, A. C., Archbald, D. A., & Tytree, A. K. Jr. (1991). Reforming the curriculum: Will empowerment policies replace control? In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing: The 1990 yearbook of the Politics of Education Association* (pp. 11-36). Bristol, PA: Falmer Press.

Porter, A. C., Floden, R., & Fuhrman, S. (1998). The effects of upgrading policies on high school mathematics and science. *Brookings Papers on Education Policy,* , 123-172.

Porter, A. C., McMaken, J., Hwang, J., & Yang, R. (2011a). Assessing Common Core Standards: Opportunities for improving measures of instruction. *Educational Researcher*, *40*(4), 186-188.

Porter, A. C., McMaken, J., Hwang, J., & Yang, R. (2011b). Common Core Standards: The new U.S. intended curriculum. *Educational Researcher*, *40*(3), 103-116.

Porter, A. C., Polikoff, M. S., & Smithson, J. (2009). Is there a de facto national intended curriculum? Evidence from state standards. *Educational Evaluation and Policy Analysis, 31*(3), 238-268.

Porter. A. C., & Smithson, J. L. (2001). *Defining, developing, and using curriculum indicators.* Report No. RR-048. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

Potterton, M. (2006). Standards – a couple of steps back. *Conference papers of the 4th Sub-Regional Conference on Assessment in Education* (pp. 295 -304). Pretoria, South Africa: Umalusi.

Price, M. (2010, April 21). Race a fair varsity admission criterion − for now. *Cape Times*, South Africa.

Prinsloo, J. (2004). Examining the examination: the wor(l)ding of the matriculation examinations in KwaZulu-Natal. *Perspectives in Education, 22*(1), 81-97.

Provincial Administration of the Cape of Good Hope. (1987). *Guidelines to Biology teachers when implementing the new syllabus in Std. 10* [Grade 12]. Cape Town, South Africa: Author.

Raizen, S. A. (1997). *Standards for science education.* Occasional Paper No.1. Madison, WI: National Institute for Science Education.

Raizi, A. M. (2010). Evaluation of learning objectives in Iranian high-school and pre-university English textbooks using Bloom's Taxonomy. *The Electronic Journal for English as Second Language, 13*(4). Retrieved November 22, 2011, from http://www.tesl-ej.org/wordpress/

Rakometsi, M. S. (2011). *Opening remarks at the media briefing on Umalusi's disclosure of the 2010 National Senior Certificate (NSC) standardisation decisions*. Pretoria, South Africa: Umalusi.

Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1993). Higher order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational Research Journal,* 30(3), 523-553.

Reddy, V. (2006a). Introduction. In V. Reddy (Ed.), *Marking Matric: Colloquium Proceedings* (pp. xii-xix). Cape Town: Human Sciences Research Council Press.

Reddy, V. (2006b). *Mathematics and science achievement in South African schools in TIMSS 2003.* Cape Town, South Africa: Human Sciences Research Council Press.

Resnick, L. B. (1987). *Education and learning to think*. Washington, DC: National Academy Press.

Resnick, L. B., Nolan, K. J., & Resnick, D. P. (1995). Benchmarking education standards. *Educational Evaluation and Policy Analysis, 17*(4), 438-461.

Resnick, L. B., Nolan, K. J., & Resnick, D. P. (1996). Caution heeded: A response to "High standards and cultural diversity: Cautionary tales of comparative research". *Educational Evaluation and Policy Analysis, 18*(3), 262-264.

Resnick, L. B., Rothman, R., Slattery, J. B., & Vranek, J. L. (2003). Benchmarking and alignment of standards and testing. *Educational Assessment, 9*(1&2), 1-27.

Richardson, J. (2009/2010). Playing 'catch-up' with developing nations makes no sense for U. S. *Phi Delta Kappan, 91*(4), 15-20.

Robelen, E. W. (2010). Work beings on 'next generation' of science standards. *Education Week,29*(21)*,* 8-9. Retrieved May 22, 2010, from http://www.edweek.org/ew/articles/2010/02/10/21science.h29.html

Roberts, N. (1976). Further verification of Bloom's Taxonomy. *Journal of Experimental Education, 45*(1),16-19.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 22-57.

Robinson, S. N. (2011 July 12). Book review − *Someone has to fail: The zero-sum game of public schooling* by David F. Labaree, 2010. *Education Review*, *14.* Retrieved July 12, 2011, from http://www.edrev.info/reviews/rev1101.pdf

Robitaille, D. F., McKnight, C., Schmidt, W. H., Britton, E., Raizen, S., & Nicol, C. (1993). *Curriculum frameworks for mathematics and science* The Third International Mathematics and Science Study, TIMSS Monograph No. 1. Vancouver, Canada: Pacific Educational Press.

Rodriguez, A. J. (1997). The dangerous discourse of invisibility: A critique of the National Research Council's National Science Education Standards. *Journal of Research in Science Teaching, 34*(1), 19-37.

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice, 24*(2), 3-13.

Rogers, W. T., & Dawber, T. E. (2002). Standard setting should include both the development of performance standards and the setting of cutscores. Book review of *Setting performance standards: concepts, methods and perspectives*, G. J. Cizek (Ed.), 2001. *International Journal of Testing*, *2*(2), 161-168.

Roth, W. (1993). Constructing knowledge from science laboratory activities. In J. H. Clarke & A. W. Biddle (Eds.), *Teaching critical thinking* (pp. 238-253). Englewood, NJ: Prentice Hall.

Rothman, B. (2009). *Common standards: The time is now.* Issue brief. Washington, DC: Alliance for Excellent Education. Retrieved March 30, 2011, from http://www.all4ed.org/files/OnlineLearning.pdf

Rothman, R. (2003). *Imperfect matches: The alignment of standards and tests.* Paper commissioned by the Committee on Test Design for K-12 Science Achievement, Centre for Education, National Research Council. Washington, DC: National Academy of Sciences.

Rohwer, W. D., & Sloane, K. (1994). Psychological perspectives. In L. W. Anderson, & L. A. Sosiak (Eds.), *Bloom's Taxonomy: A forty-year retrospective, Ninety-third Yearbook of the National Society for the Study of Education, Part II* (pp. 64-81). Chicago: University of Chicago Press.

Rutherford, F. J. (2003). Reflections in assessment. In J. M. Atkin, & J. E. Coffey (Eds.), *Everyday* a*ssessment in the science classroom* (pp. 147-158). Washington, DC: National Science Teachers Association Press.

Ryan, K. (2002). Assessment validation in the context of high-stakes testing assessment. *Educational Measurement: Issues and Practice, 21*(1), 7-15.

Sadler, D. R. (1996). *Criteria and standards in student assessment − Different approaches: Theory and practice in higher education.* Paper presented at the HERDSA Conference, Perth, Western Australia, 8-12 July. Retrieved March 13, 2006, from http://www.herdsa.org.au/confs/1996/sadler.html

Sax, G., & Collet, L. S. (1968). A empirical comparison of the effects of recall and multiple choice tests on student achievement. *Journal of Educational Measurement, 5*(2): 169-173.

Sax, G., Eilenberg, E. G., & Klockars, A. J. (1972) Achievement as a function of test item complexity and difficulty. *Journal of Experimental Education, 40*(4), 90-93.

Sax, G., & Reade, M. (1964). Achievement as a function of test difficulty level. *American Educational Research Journal,* 1(1), 22-25.

Schafer, W. D. (2011). Aligned by design: A process for systematic alignment of assessments to educational domains. In G. Schraw & D. R. Robinson (Eds.), *Assessment of higher order thinking skills* (pp. 395-418). Charlotte, NC: Information Age Publishing.

Scharaschkin, A., & Baird, J. (2000). The effects of consistency of performance on A Level examiners. judgements of standards. *British Educational Research Journal*, *26*(3), 343-357.

Scheltema, A., & Myburgh, C. (1990). *Bios '90 H*[igher]*G*[rade] *Senior Certificate Biology examination pap*ers [1989] *with answers for Std 10* [Grade 12] *pupils*. Stellenbosch, South Africa: Fisichem Publishers.

Scheltema, A., & Myburgh, C. (1995). *Bios '95 H*[igher]*G*[rade] *(Eng) Senior Certificate Biology examination papers* [1994] *with answers for Std 10* [Grade 12*] pupils*. Stellenbosch, South Africa: Fisichem Publishers.

Scheltema, A., & Myburgh, C. (1996). *Bios '96 H*[igher]*G*[rade] *Senior Certificate Biology examination papers* [1995] *with answers for Std 10* [Grade 12] *pupils*. Stellenbosch, South Africa: Fisichem Publishers.

Scheltema, A., & Myburgh, C. (1997*). Bios '97 H*[igher]*G*[rade] *Senior Certificate Biology examination papers* [1996] *with answers for Std 10* [Grade 12] *pupils*. Stellenbosch, South Africa: Fisichem Publishers.

Scheltema, A., & Myburgh, C. (1998). *Bios '98 H[*igher]*G*[rade] *Senior Certificate Biology examination papers* [1997] *with answers for Gr*[a]*d*[e] *12 pupils*. Stellenbosch, South Africa: Fisichem Publishers.

Scheltema, A., & Myburgh, C. (1999). *Bios '99 H*[igher]*G*[rade*] Senior Certificate Biology examination papers* [1998] *with answers for Gr*[a]*d*[e] *12 pupils*. Stellenbosch, South Africa: Fisichem Publishers.

Scheltema, A., & Myburgh, C. (2000). *Bios 2000 H*[igher]*G*[rade] *Senior Certificate Biology examination papers* [1999] *with complete answers for* [G]*rade 12 pupils*. Stellenbosch, South Africa: Fisichem Publishers.

Scheltema, A., & Myburgh, C. (2001). *Bios 2001 H*[igher]*G*[rade] *Senior Certificate Biology examination papers* [2000] *with complete answers for* [G]*rade 12 pupils.* Stellenbosch, South Africa: Fisichem Publishers.

Scheuneman, J., Gerritz, K., & Embretson, S. (1991). *Effects of prose complexity on achievement test item difficulty.* Report No. ETS-RR-91-43. Princeton, NJ: Educational Testing Service.

Schraw, G., & Robinson, D. R. (2011) Conceptualizing and assessing higher order thinking skills. In G. Schraw & D. R. Robinson (Eds.), *Assessment of higher order thinking skills* (pp. 1-15). Charlotte, NC: Information Age Publishing.

Schwille, J. (1996). Commentary [−] Understanding the examination systems. In E. D. Britton & S. A. Raizen (Eds.), *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students* (pp. 199-200). Boston: Kluwer Academic Publishers.

Schmidt, W. H., Houang, R. T., & Cogan, L. (2002). A coherent curriculum: The case of mathematics. *American Educator, 26*(2), 1-18.

Schmidt, W. H., & McKnight, C. C. (1995). Surveying educational opportunity in Mathematics and Science: An international perspective. *Educational Evaluation and Policy Analysis, 17*(3), 337-353.

Schmidt, W. H., McKnight, C. C., Valverde, G., Houang, R. T., & Wiley, D. E. (1997). *Many visions, many aims - Volume 1: A cross-national investigation of curriculum intentions in school mathematics.* Dordrecht: Kluver Academic Publishers.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. C., Wiley, D. E, & Cogan, L. S. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.

Schmidt, W. H., Wang, H. C., & McKnight, C. M. (2005). Curriculum coherence: an examination of US mathematics and science content standards from an international perspective. *Journal of Curriculum Studies*, *37*(5), 525-559.

Schönborn, K. J., & Bögeholz, S. (2009). Knowledge transfer in biology and translation across external representations: Experts' views and challenges for learning. *International Journal of Science and Mathematics Education, 7*(5), 931-955.

Schunn, C. (2009). *Are 21st century skills found in science standards?* Paper prepared for the Workshop on Exploring the Intersection of Science Education and the Development of 21st Century Skills, National Research Council. Retrieved June 22, 2010, from http://www.nationalacademies.org/bose/Schunn.pdf

Schwartz, M. S., Sadler, P. M., Sonnert, G., & Tai, R. H. (2009). Depth versus breadth: How content coverage in high school science courses relates to later success in college science coursework. *Science Education, 93*(5), 798-826.

Schwille, J. (1996). Commentary [−] Understanding the examination systems. In E. D. Britton & S. A. Raizen (Eds.), *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students (*pp. 199-200). Boston: Kluwer Academic Publishers.

Sebatane, M. (2000). International transfers of assessment: Recent trends and strategies. *Assessment in Education, 7*(3): 401-416.

Seddon, G. M. (1978). The properties of Bloom's Taxonomy of Educational Objectives for the cognitive domain. *Review of Educational Research, 48* (2), 303-323.

Selod, H., & Zenou, Y. (2003). Private versus public schools in post-apartheid South African cities: theory and policy implications. *Journal of Development Economics*, *71*(2), 351-394.

Seekings, J. (2002). *Indicators of performance in South Africa's public school system.* CSSR Working Paper No. 10. Cape Town, South Africa: Center for Social Science Research, University of Cape Town. Retrieved August 5, 2011, from http://www.uct.ac.za/depts/cssr

Senechal, D. (2010). The most daring education reform of all. *American Educator*, *34*(1), 4-16.

Shavelson, R. J., & Huang, L. (2003). Responding responsibly to the frenzy to assess learning in higher education. *Change, 35*(1), 10-19.

Shaw, J. M., Bunch, G. C., & Geaney, E. R. (2010), Analyzing language demands facing English learners on science performance assessments: The SALD framework. *Journal of Research in Science Teaching, 47*(8)*,* 909-928.

Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405-450.

Shepard, L. A. (2003). Reconsidering large-scale assessment to heighten its relevance to learning. In J. M. Atkin, J.M & J. E. Coffey (Eds.), *Everyday* assessment *in the science classroom* (pp. 121-146). Washington, DC: National Science Teachers Association Press.

Shepard, L. A. (2010). What the marketplace has brought us: Item-by-item teaching with little instructional insight. *Peabody Journal of Education*, *85*(2), 246–257.

Shulman, L. S. (2002). Making differences: A table of learning. *Change, 34*(6), 36-44.

Shulman, L. S. (2007). Counting and recounting: Assessment and the quest for accountability. *Change, 39*(1), 20-25.

Silva, E. (2009). Measuring skills for 21st-century learning. *Phi Delta Kappan, 90*(9), 630-634.

Simkins, C. (2010). *The maths and science performance of South Africa's public schools ─ Some lessons from the past decade.* Johannesburg, South Africa: The Centre for Development and Enterprise.

Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, *45*(1-3), 83-117.

Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher, 36*(8), 477-481.

Sireci, S. G. (2009). Packing and unpacking sources of validity evidence: History repeats itself again. In R.W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 19-37). Charlotte, NC: Information Age Publishing.

Sireci, S. G., & Parker, P. (2006). Validity on trial: Psychometric and legal conceptualizations of validity. *Educational Measurement, Issues and Practices, 25*(3), 27-33.

Smagorinsky, P. (1998). Thinking and speech protocol analysis. *Mind, Culture, and Activity. 5*(3), 157-177.

Smith, I. L. (1970). I. Q., creativity, and the Taxonomy of Educational Objectives: Cognitive domain. *Journal of Experimental Education, 38*(4), 58-60.

Smith, R. B. (1968). An empirical examination of the assumptions underlying the "Taxonomy of Educational Objectives: Cognitive domain". *Journal of Educational Measurement, 5*(2): 125-128.

Smith, R. B. (1970). An empirical investigation of complexity and process in multiple-choice items. *Journal of Educational Measurement, 7*(1): 33-41.

Snow, R. E. (1994). *Abilities in academic tasks*. In R. J.Sternberg & R. K.Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 3–37). Cambridge, England.

Sousa, D. A. (2011). *How the ELL brain learns.* Thousand Oaks, CA: Corwin Press.

South African Institute of Race Relations. (2008). *South African Survey 2005/2006.* Johannesburg, South Africa: Author.

Spencer, A. (2003). Changes in the nature of assessment. In *Setting the standard* (pp. 115-136). Manchester, England: The Assessment and Qualifications Alliance.

Stanley, G., & Tognolini, J. (2008). *Performance with respect to standards in public examinations*. Paper presented at the International Association for Educational Assessment 34th Annual Conference, Cambridge, UK. Retrieved November 3, 2010, from http://iaea2007.tqdk.gov.az/cp/Techniques%20for%20monitoring%20the%20comparability%20of%20examination%20standards.pdf

Stedman, C. H. (1973). An analysis of the assumptions underlying the Taxonomy of Educational Objectives: Cognitive domain. *Journal of Research in Science Teaching, 10*(3): 235-241.

Steen, L. A. (1999). Assessing assessment. In B. Gold, S. Keith & W. Marion (Eds.), *Assessment Practices in Undergraduate Mathematics.* Washington, DC: Mathematical Association of America. Retrieved March 21, 2007, from http://www.stolaf.edu/people/steen/Papers/assessment.html

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR Learning. *Phi Delta Kappan. 83*(10), 758-765. Retrieved January 8, 2006, from http://www.pdkintl.org/kappan/k0206sti.htm

Stoker, H. W., & Kropp, R. P. (1964). Measurement of cognitive processes. *Journal of Educational Measurement*, 1(1), 39-42.

Strachan, J. (2002). Assessment in change: Some reflections on the local and international background to the National Certificate of Educational Achievement (NCEA). *New Zealand Annual Review of Education*, *11*, 245- 262.

Swanson, C. B. (2006). *Making the Connection - A decade of standards-based reform and achievement.* Washington, DC: Editorial Projects in Education Research Centre. Retrieved January 7, 2006, from http://www.edweek.org/ew/articles/2006/01/05/17overviews2.h25.html?qs=Making+the+connection

Tamir, P. (1974). An inquiry orientated laboratory examination. *Journal of Educational Measurement, 11*(1), 25-33.

Tamir, P. (1996). Biology examinations. In E. D. Britton & S. A. Raizen (Eds.), *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students* (pp. 55-85). Boston: Kluwer Academic Publishers.

Taylor, N. (2009, January 28). Raise the standards when examining matric results. *Business Day,* South Africa.

Taylor, N., Muller, J., & Vinjevold, P. (2003). *Getting schools working – research and systemic school reform in South Africa.* Cape Town, South Africa: Maskew Miller Longman.

The Childrens Institute. (2010). *2008/2009 South African child gauge* [:] *Children and education*. Cape Town, South Africa: University of Cape Town.

Thomas B. Fordham Foundation (2005, December 7). Press release: *Most K-12 state science standards don't make the grade.* Retrieved December 14, 2005, from http://www.edexcellence.net/foundation/about/press_release.cfm

Tindal, G., & Almond, P. (2006). *Alignment of alternate assessments with state content standards.* Retrieved May 3 2007 from http://www.daata.org/media/content/ContEvidResearch103006.pdf

Tognolini, J., & Stanley, G. (2007). Standards-based assessment: a tool and means to the development of human capital and capacity building in education. *Australian Journal of Education, 51*(2), 129-145.

Towne, L., Wise, L., & Winter, T. (Eds.) 2004. *Advancing scientific research in education.* Washington, DC: National Academies Press.

Travers, K. J., Crosswhite, F. J., Dossey, J. A., Swafford, J. O., McKnight, C. C., & Cooney, T. J. (1985). *Second International Mathematics Study summary report for the United States.* Champaign, IL: Stipes Publishing Company.

Trümpelmann, M. H. (1991). *The Joint Matriculation Board: Seventy-five years achievement in perspective* (P. R. T. Nel, Trans.). Cape Town, South Africa: National Book Printers.

Tyler, R.W. (1949). *Basic principles of curriculum and instruction.* Chicago, IL: University of Chicago Press.

Umalusi. (2004). *Investigation into the standard of the Senior Certificate Examination.* Research report. Pretoria, South Africa: Author.

Umalusi. (2005). *Interim directives for certification by Umalusi*. Pretoria, South Africa: Author.

Umalusi. (2007). *A comparative study of the intended curriculum and exit-level examinations in biology in four African countries.* Pretoria, South Africa: Author.

Umalusi. (2009a). *From NATED 550 to the new National Curriculum: maintaining standards in 2008* [–] *2008 Maintaining Standards Report – Part 1: Overview.* Pretoria, South Africa: Author.

Umalusi. (2009b). *From NATED 550 to the new National Curriculum: maintaining standards in 2008* [–] *2008 Maintaining Standards Report – Part 2: Curriculum evaluation.* Pretoria, South Africa: Author.

Umalusi. (2010). *Summary exam analysis report: Life Sciences*. Pretoria, South Africa: Author.

Umalusi. (2011, February 23). *How Umalusi adjusted the 2010 matric results.* A text of briefing by Chairperson Sizwe Mabizela. Pretoria, South Africa: Author. Retrieved July 22, 2011 from http://www.politicsweb.co.za/politicsweb/view/politicsweb/en/page71656Detail

University of Cape Town. (2001). *A guideline for acceptance into UCT for 2002*. Cape Town, South Africa: Author.

University of Cape Town. (2002). *A guideline for acceptance into UCT for 2003*. Cape Town, South Africa: Author.

University of Cape Town. (2003). *A guideline for acceptance into UCT for 2004.* Cape Town, South Africa: Author.

University of Cape Town. (2004). *A guideline for acceptance into UCT for 2005.* Cape Town, South Africa: Author.

University of Cape Town. (2005). *A guideline for acceptance into UCT for 2006.* Cape Town, South Africa: Author.

University of Cape Town. (2006). *A guideline for acceptance into UCT for 2007.* Cape Town, South Africa: Author.

Valverde, G. A. (2000). Strategic themes in curriculum policy documents: An exploration of TIMSS curriculum analysis data. *International Journal of Educational Policy, Research and Practice*, *1*(2), 133-152.

Valverde, G. A. (2004). Curriculum convergence in Chile: The global and local context of reforms in curriculum policy. *Comparative Education Review*, *48*(2), 174–201.

Valverde, G. A. (2005). Curriculum policy seen through high-stakes examinations: Mathematics and Biology in a selection of school leaving examinations from the Middle East and North Africa. *Peabody Journal of Education*, *80*(1), 29-55.

Valverde, G. A., & Schmidt, W. H. (2000). Greater expectations: learning from other nations in the quest for 'world-class' standards in US school mathematics and science. *Journal of Curriculum Studies*, *32*(5), 651-687.

Van Dantzig, D. (1978). Data analysis, exploratory. In W. H. Kruskal & J. M. Tanur (Eds.), *International encyclopedia of statistics – Volume 1* (pp. 97- 336). New York: The Free Press.

Vandeyar, S., & Killen, R. (2003). Has curriculum reform in South Africa really changed assessment practices, and what promise does the revised National Curriculum Statement hold? *Perspectives in Education, 21*(1), 119-134.

Van der Berg, S., & Burger, R. (2003). Education and socio-economic differentials: A study of school performance in the Western Cape. *South African Journal of Economics, 71*(3): 496–522.

Venter, C., & Cohen, D. (2004). The century of biology. *New Perspectives Quarterly, 21*(4): 73-77.

Vitale, M. R., & Romance, N. R. (2006). Research in science education: An interdisciplinary perspective. J. Rhoton & P. Shane (Eds.), *Teaching science in the 21st century* (pp. 329-351). Arlington, VA: National Science Teachers Association Press.

Vitale, M. R., Romance, N. R., & Dolan, M. F. (2006). A knowledge-based framework for the classroom assessment of student science understanding. In M. McMahon, P. Simmons, R. Sommers, D. DeBaets & F. Crawley (Eds.), *Assessment in science – Practical experiences and education research* (pp. 1-14). Arlington, VA: National Science Teachers Association Press.

Wagner, R. K. (2003). Sketching a framework for assessment: prospects for moving beyond principles to practice. *Measurement: Interdisciplinary Research and Perspective*, *1*(1), 88-91.

Wainer, H., & Thissen, D. (1993). Combining multiple choice and constructed response test scores: Towards a Marxist theory of test construction. *Applied Measurement in Education, 6*(2), 103-118.

wa Kivilu, M. (2006). The matriculation examination: How can we find out if standards are falling? In V. Reddy (Ed.), *Marking Matric: Colloquium Proceedings* (pp. 33-44). Pretoria: Human Sciences Research Council Press.

Wang, V., & Farmer, L. (2008). Adult teaching methods in china and Bloom's Taxonomy. *International Journal for the Scholarship of Teaching and Learning, 2*(2), 1-16.

Webb, N. L. (1997a). *Criteria for alignment expectations and assessments in mathematics and science education.* Research Monograph No. 8. Washington, DC: Council of Chief State School Officers.

Webb, N. L. (1997b). Determining alignment of expectations and assessments in mathematics and science education. *National Institute of Science Education Brief, 1*(2), 1-10.

Webb, N. L. (1999). *Alignment of science and mathematics standards in four states.* Research Monograph No.18. Washington, DC: Council of Chief State School Officers.

Webb, N. L. (2002). *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states.* Washington, DC: Council of Chief State School Officers.

Webb, N. L. (2006). Identifying content for student achievement tests. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 155-180). Mahwah, NJ: New Jersey: Lawrence Erlbaum Associates.

Webb, N.L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, *20* (1), 7-25.

Welch, C. (2006). Item and prompt development in performance testing. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 303-328). Mahwah, NJ: Lawrence Erlbaum Associates.

Wendler, C. L. W., & Walker, M. E. (2006). Practical issues in designing and maintaining multiple test forms for large-scale programs. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 445-467). Mahwah, NJ: Lawrence Erlbaum Associates.

Western Cape Education Department. (1995a). *Instructional Offering: Biology – Instructional Programmes for Higher Grade: Standards 8, 9 and 10* [Grades 10,11,12]. Cape Town, South Africa: Author.

Western Cape Education Department. (1995b). *Instructional Offering: Biology – Instructional Programmes for Standard Grade: Standards 8, 9 and 10* [Grades 10,11,12]. Cape Town, South Africa: Author.

Western Cape Education Department. (1996). *Guidelines: Biology HG/SG:* [Senior Certificate] *Examination 1996/97*. Cape Town, South Africa: Author.

Whiteley, T. R. (2006). Using the Socratic Method and Bloom's Taxonomy of cognitive demand to enhance online discussion, critical thinking, and student learning. *Developments in Business Simulation and Experiential Learning, 33,* 65-70. Retrieved June 22, 2008 from http://sbaweb.wayne.edu/~absel/bkl/.%5Cvol33%5C33ai.pdf

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research & Evaluation*, 2(2). Retrieved March 17, 2007, from http://pareonline.net/getvn.asp?v=2&n=2

Wiggins, G. (1991a). *Toward assessment worthy of the liberal arts: The truth may make you free, but the test may keep you imprisoned.* Appendix to Mathematics Association of America (2002): Getting Started With Assessment. *MAA Notes, 49*. Retrieved March 21, 2007, from http://www.maa.org/saum/gettingstarted.html

Wiggins, G. (1991b). Standards, not standardization: Evoking quality student work**.** *Educational Leadership, 48*(5): 18-25.

Wiggins, G. (1997). Work standards: Why we need standards for instructional and assessment design. *National Association of Secondary School Principals Bulletin*, 81(590), 56-64.

Wiggins, G. (2011). A diploma worth having. *Educational Leadership*, *68*(6), 28–33.

Wiggins, G., & McTighe, J. (2008). Put understanding first. *Educational Leadership, 65*(8), 36-41.

Wiliam, D. (1996a). Standard-setting methods for multiple levels of competence. In B. Boyle & T. Christie (Eds.), *Issues in setting standards: establishing comparabilities* (pp.71-83). London: Falmer Press.

Wiliam, D. (1996b). Standards in examinations: A matter of trust? *The Curriculum Journal, 7*(3), 293-306.

Williamson, P. (2003). Setting, marking and awarding: The examination process. In *Setting the standard* (pp. 161-170). Manchester, England: The Assessment and Qualifications Alliance.

Willingham, D. (2007). Critical thinking: Why is it so hard to teach? *American Educator, 31*(2), 8-19.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement, 39*(1), 1-37.

Willson, I. A. (1973). Changes in mean levels of thinking in Grades 1-8 through use of an interaction analysis system based on Bloom's Taxonomy. *Journal of Educational Research, 66*(9), 423-429.

Wineburg, S. (1997). Beyond "breadth and depth": Subject matter knowledge and assessment. *Theory into Practice, 36*(4), 255-261.

Wineburg, S., & Schneider, J. (2009/2010). Was Bloom's Taxonomy pointed in the wrong direction? *Phi Delta Kappan, 91*(4), 56-61.

Wiske, M. S. (1998). What is teaching for understanding? In M. S. Wiske (Ed.), *Teaching for understanding: Linking research with practice* (pp. 61-86). San Francisco: Jossey-Bass Publishers.

Wood, R. (1977). Multiple choice: A state of the art report. *Evaluation in Education: International Progress, 1*, 191-280.

Wurtz, E., & Malcom, S. (1993). *Promises to keep: Creating high standards for American students*. Report on the Review of Education Standards from the Goals 3 and 4 Technical Planning Group to the National [USA] Education Goals Panel. Retrieved January 10, 2009, from http://govinfo.library.unt.edu/negp/reports/promises.pdf

Yager, R. E. (1993). Science and critical thinking. In J. H. Clarke & A. W. Biddle (Eds.), *Teaching critical thinking* (pp. 264-275). Englewood, NJ: Prentice Hall.

Yager, R. E., & McCormack, A. J. (1989). Assessing teaching/learning successes on mjultiple domains of science and science education. *Science Education, 73*(1), 45-58.

Yeld, N. (2001). *Assessment, equity and language of learning: Key issues for higher education selection in South Africa.* Doctoral dissertation, University of Cape Town, Cape Town, South Africa.

Yeld, N. (2005a, September 9). Failed by stale tests. *Financial Mail,* South Africa.

Yeld, N. (2005b). Senior Certificate syllabus stagnation and attendant woes: Some consequences for National Senior Certificate. In *Matric: What is to be done?* (pp. 45-57). Seminar papers and presentations – Umalusi and the Centre for Higher Education Transformation Seminar on Matric. Pretoria, South Africa: Umalusi.

Yeld, N. (2009). The National Benchmarks Tests Project: Addressing student educational needs in the tertiary education system. In: J.Hofmeyr (Ed.), *Education and Skills 2009 Transformation Audit* (pp.76-83). Cape Town, South Africa: The Institute for Justice and Reconciliation.

Young, J. W., Cho, Y., Ling, G., Cline, F., Steinberg, J., & Stone, E. (2008). Validity and fairness of states standards-based assessments for English language learners. *Educational Assessment,13*(2/3), 170-192.

Young, J. W., Steinberg, J., Cline, F., Stone, E., Martiniello, M., Ling, G., & Cho, Y. (2010). Examining validity of standards-based assessments for initially fluent students and former English language learners. *Educational Assessment, 15*(2), 87-106.

Zheng, A. J., Lawhorn, J. K., Lumbley, T., & Freeman, S. (2008). Applications of Bloom's Taxonomy debunks the "MCAT Myth". *Science, 319,* 414-415.

Zieky, M. (2006). Fairness reviews in assessment. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 359-378). Mahwah, NJ: Lawrence Erlbaum Associates.

Zietsman, A., & Gering, M. (1986). Admission to university in an academically non-homogenous society. *Higher Education, 15*(1/2), 25-35.

Zille, H. (2005, May 22). Government has failed to improve education system. *The Sunday Independent*, South Africa.

Zohar, A. & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *The Journal of the Learning Sciences*, *12*(2): 145-181.

Zohar, A., Schwartzer, N., & Tamir, P. (1998). Assessing the cognitive demands required of students in class discourse, homework assignments and tests. *International Journal of Science Education, 20*(7), 769-782.

Zoller, U. (1993). Are lecture and learning compatible? *Journal of Chemical Education, 70*(3), 195-197.

Zoller, U., & Tsaparlis, G. (1997). Higher[-] and lower-order cognitive skills: The case of chemistry. *Research in Science Education, 27*(1), 117-130.

Zucker, S., Sassman, C., & Case, B. J. (2004). *Cognitive labs.* Technical report. San Antonio, TX: Harcourt Assessment. Retrieved on March 7, 2007, from http://harcourtassessement.cm/hai/Images/resource/library/pdf/CognitiveLabsFinal.pdf

Zull, J. E. (2002). *The art of changing the brain: Enriching the practice of teaching by exploring the biology of learning.* Sterling, VA: Stylus Publishing.

Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validity practice. In R.W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 65-82). Charlotte, NC: Information Age Publishing.

STANDARDS OF SOUTH AFRICAN SENIOR CERTIFICATE BIOLOGY EXAMINATIONS: 1994 TO 2007

VOLUME II: Appendices

Anna Aletta Crowe

Thesis presented for the degree of

DOCTOR OF PHILOSOPHY

in the School of Education

UNIVERSITY OF CAPE TOWN

July, 2012

Supervisors: Associate Professor Rüdiger Laugksch
Professor Timothy Dunne

APPENDICES

Appendix 2.1 An example of the principles of marking attached to each marking memorandum (DoE, 2004d).

PRINCIPLES RELATED TO MARKING HG & SG BIOLOGY 2004

This document should be attached to all memoranda; attached to all updated guidelines that are distributed in 2005 and made available to ALL Biology teachers early in 2005.

1. **If more information than marks allocated is given**
   Stop marking when maximum marks is reached and put a wavy line and ómaxô in the right hand margin.

2. **If, for example, three reasons are required and five are given**
   Mark the first three irrespective of whether all or some are correct/incorrect.

3. **If whole process is given when only part of it is required**
   Read all and credit relevant part.

4. **If comparisons are asked for and descriptions are given**
   Accept if differences / similarities are clear.

5. **If tabulation is required but paragraphs are given**
   Candidates will lose marks for not tabulating.

6. **If diagrams are given with annotations when descriptions are required**
   Candidates will lose marks

7. **If flow charts are given instead of descriptions**
   Candidates will lose marks

8. **If sequence is muddled and links do not make sense**
   Where sequence and links are correct, credit. Where sequence and links are incorrect, do not credit. If sequence and links become correct again, resume credit.

9. **Non-recognized abbreviations**
   Accept if first defined in answer. If not defined, do not credit the unrecognised abbreviation but credit the rest of answer if correct.

10. **Wrong numbering**
    If answer fits into the correct sequence of questions but the wrong number is given, it is acceptable.

11. **If language used changes the intended meaning**
    Do not accept.

12. **Spelling errors**
    If recognizable accept provided it does not mean something else in Biology or if it is out of context.

13. **If common names given in terminology**
    Accept provided it is accepted at *this* memo discussion.

14. **If only letter is asked for and only name is given (and vice versa)**
    No credit

15. **If units are not given in measurements**
    Candidates will lose marks. Memorandum will allocate marks for units separately. (*Added at the end of 2009:* except where it is already given in the question.)

16. **Be sensitive to the sense of an answer, which may be stated in a different way**.

17. **Caption**
    All illustrations (diagrams, graphs, tables, etc.) must have a caption. (*Added at the end of 2009:* except where it is already given in the question.)

18. If you have doubts consult the other language memo, if you still have doubts ask the Provincial Internal Moderator to contact the National Internal Moderator or the External Moderators.

19. **Code switching of official languages (terms and concepts)**
    A single word or two that appears in any official language other than the learners assessment language used to the greatest extent in his or her answers should be credited, if it is correct. A marker that is proficient in the relevant official language should be consulted. This is applicable to all official languages.

20. No changes must be made to the marking memoranda without consulting the Provincial Internal Moderator who in turn will consult with the External Moderator/s

Appendix 2.1 continued

Added at the end of 2006

21. Only memoranda bearing the signatures of the UMALUSI moderators and distributed by the National Department of Education via the Provinces must be used.

Appendix 4.1 Summaries of other taxonomies of cognitive demand considered, and rejected, for use in this study.

A. Bloom's Taxonomy of educational objectives used by Zohar et al. (1998).

Knowledge – as defined in Bloom et al. 1956	**Comprehension** – as defined in Bloom et al. 1956	**Lower level application** – application of scientific principles to solve quantitative problems, (manipulating equipment, performing laboratory procedures) [a]	**Higher order thinking (HOTS)** – higher level, application, analysis, synthesis, evaluation

Note:

a Zohar et al. (1998, p. 770) do not classify these categories of performance expectations as lower level application but state that they "are not higher-order thinking skills"

B. An extrapolation of Bloom's Taxonomy of educational objectives (Crowe, Dirks & Wenderoth, 2008).

Knowledge – **identify, recall**, list recognize, or label	**Comprehension** – **describe** or explain in your own words, re-tell, or summarize	**Application** – **predict** an outcome using several pieces of information or concepts; use information in a new context	**Analysis** – **infer**; understand how components relate to each other and to the process as a whole	**Synthesis** – **create** something new using / combining disparate sources of information	**Evaluation** – **determine / critique** relative value; determine merit

C. The types of knowledge and use of types knowledge for science assessments (Li, 2001).

Declarative	**Procedural**	**Schematic**	**Strategic**
Defining, comparing/ contrasting, exemplifying and explaining	Executing and performing	Explaining, justifying and predicting/ hypothesizing	Framing patterns, raising questions and defining problems, choosing types of knowledge and assembling knowledge for use

D. The levels of student understanding recognized in the SOLO taxonomy (Biggs,1999).

Unistructural	**Multistructural**	**Relational**	**Extended abstract**
1. Identify *2. Do simple procedure*	*1. Enumerate* *2. Describe* *3. List* *4. Combine* *5. Do algorithms*	*1. Compare / contrast* *2. Explain causes* *3. Analysis* *4. Relate* *5. Apply*	*1. Theorize* *2. Generalize* *3. Hypothesize* *4. Reflect*

E. The Coding Procedures for Curriculum Content Analyses – cognitive demand categories for science (CCSS0 & WCER, 2004).

Memorize	**Perform procedures**	**Communicate understanding**	**Analyze information**	**Apply concepts / make connections**

Appendix 4.1 continued

F. TIMSS Performance Expectations: detailed categories used in TIMSS 1995 (Robitaille et al., 1993) and TIMSS-R 1999 (Martin et al., 2000).

Understanding	Theorizing, analyzing and solving problems	Using tools, routine procedures and scientific processes	Investigating the natural world	Communicating [a]
1. *Simple information* 2. *Complex information* 3. *Thematic information* [a]	1. *Abstracting and deducing scientific principles* 2. *Applying scientific principles to solve quantitative problems* 3. *Applying scientific principles to develop explanations* 4. *Constructing, interpreting and applying models* 5. *Making decisions* [a]	1. *Using apparatus, equipment and computers* 2. *Conducting routine experimental operations* 3. *Gathering data* 4. *Organizing and representing data* 5. *Interpreting data*	1. *Identifying questions to investigate* [a] 2. *Designing investigations* 3. *Conducting investigations* 4. *Interpreting investigational data* 5. *Formulating conclusions from investigational data*	1. *Accessing and processing information* 2. *Sharing information*

Note:

a Categories not used in the Britton and Raizen (1996) comparison of international science examinations for college-bound students.

G. The cognitive domain used in TIMSS 2003 analyses (Reddy, 2006). The cognitive domain replaces performance expectations used in the TIMSS 1995 and TIMSS-R 1999.

Factual knowledge	Conceptual understanding	Reasoning and analysis
1. *Recall/recognize* 2. *Define* 3. *Describe* 4. *Use tools and procedures*	1. *Illustrate with examples* 2. *Compare/contrast/ classify* 3. *Represent/model* 4. *Relate* 5. *Extract/apply information* 6. *Find solutions* 7. *Explain*	1. *Analyze/interpret/solve problems* 2. *Integrate/synthesize* 3. *Hypothesize/predict* 4. *Design/plan* 5. *Collect/analyze/interpret data* 6. *Draw conclusions*

H. Webb's Depth of Knowledge categories (Webb, 1999).

Recall	Skill/concept	Strategic thinking	Extended thinking
Recall of fact, information or procedure	Use of information, conceptual knowledge, procedures; two or more steps etc.	Requires reasoning, developing a plan or sequence of steps; has some complexity; more than one possible answer; generally takes less than 10 mins to do	Requires an investigation; time to think and process multiple conditions of the problem or task; and more than 10 mins to do non-routine manipulations

I. Marzano's New Taxonomy of educational objectives (Marzano, 2001). Changes in Marzano and Kendall (2007) are show in brackets in the table.

<table>
<tr><th rowspan="3">Knowledge domains</th><th colspan="6">Systems of thinking</th></tr>
<tr><th colspan="4">Cognitive</th><th rowspan="2">Metacognitive – monitoring, evaluating and regulating the functioning of all others through
1. Goal specification (specifying goals)
2. Process monitoring
3. Monitoring clarity
4. Monitoring accuracy</th><th rowspan="2">Self-system – interrelated system of attitudes, beliefs and emotions that determine motivation and attention
1. Examining importance
2. Examining efficacy
3. Examining emotional response
4. Examining motivation</th></tr>
<tr><th>Retrieval – the activation and transfer of knowledge from permanent memory to working memory
1. Recall (recalling)
2. Execution (executing)
(3. Recognizing)</th><th>Comprehension – distilling knowledge down to its key characteristics
1. Synthesis (integrating)
2. Representation (symbolizing)</th><th>Analysis – elaboration of knowledge as comprehended
1. Matching
2. Classifying
3. Error analysis (analyzing errors)
4. Generalizing
5. Specifying</th><th>Knowledge utilization-
1. Decision making
2. Problem solving
3. Experimental inquiry (experimenting)
4. Investigation (investigating)</th></tr>
<tr><td>Information ï declarative knowledge
1. Details
2. Organizing ideas</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Mental procedures ï Procedural knowledge
1. Skills
2. Processes</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Psychomotor procedures ï physical procedures to engage in physical activities
1. Skills -
2. Processes</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Appendix 4.1 continued

J. The Revised BTEO (Anderson et al., 2001).

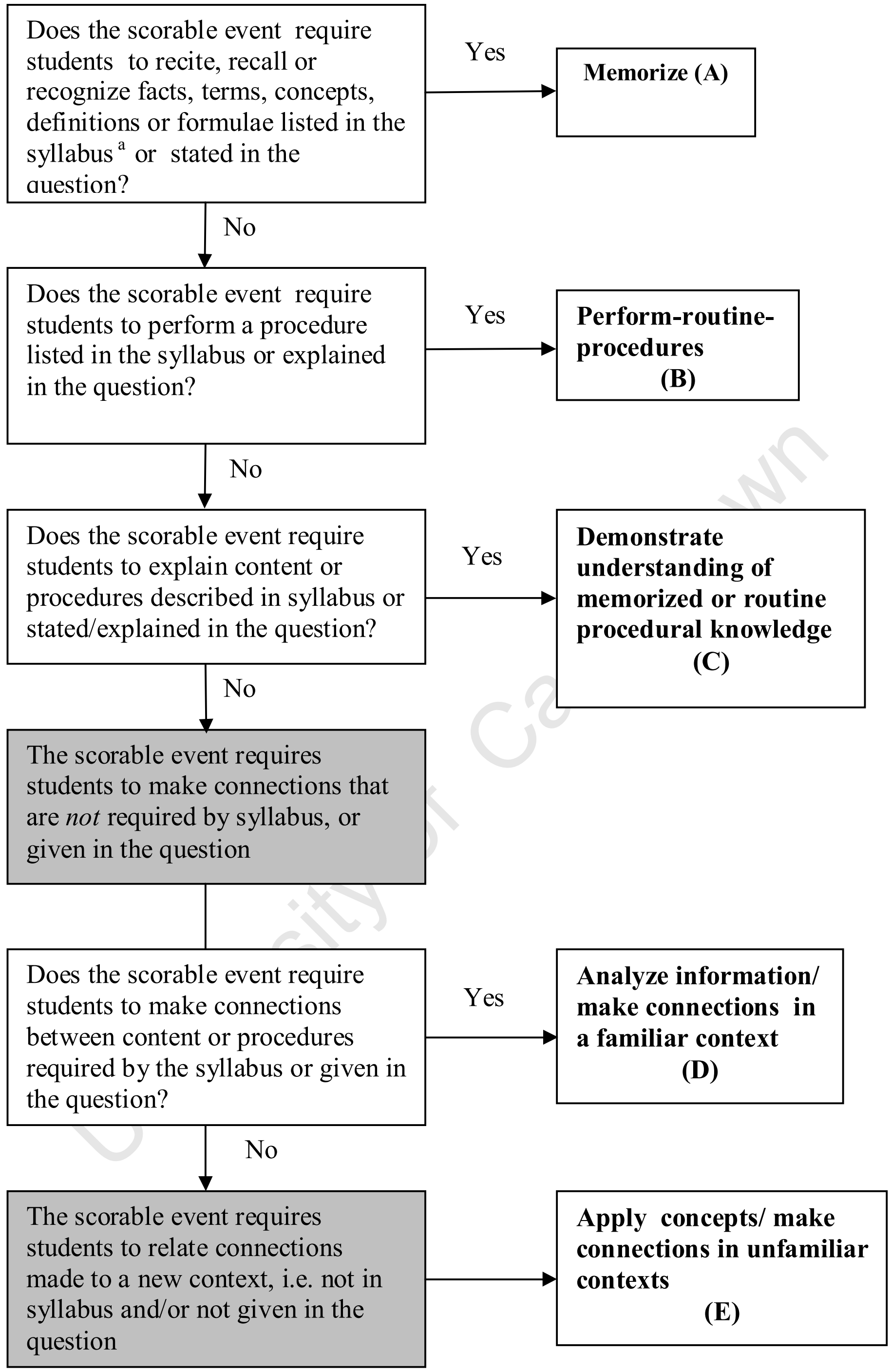
The Knowledge Dimension	**The Cognitive Process Dimension**					
	Remember - retrieving relevant knowledge from long-term memory *1. Recognizing* *2. Recalling*	**Understand -** determining meaning from instructional messages *1. Interpreting* *2. Exemplifying* *3. Classifying* *4. Summarizing* *5. Inferring* *6. Comparing* *7. Explaining*	**Apply -** carrying out a procedure in a given situation *1. Executing* *2. Implementing*	**Analyse** ï breaking material into constituent parts and determining how the parts relate to one another and to an overall structure or purpose *1. Differentiating* *2. Organizing* *3. Attributing*	**Evaluate** ï making judgments based on criteria and standards *1. Checking* *2. Critiquing*	**Create** ï putting elements together to form a coherent whole or new whole *1. Generating* *2. Planning* *3. Producing*
Factual Knowledge ï basic elements that students must know to be acquainted with a discipline or solve a problem in it *1. Knowledge of terminology* *2. Knowledge of specific details and elements*						
Conceptual Knowledge ï the interrelationships among the basic elements within a larger structure that enable them to function together *1. Knowledge of classifications and categories* *2. Knowledge of principles and generalizations* *3. Knowledge of theories, models and structures*						
Procedural Knowledge ï how to do something, methods of inquiry, and criteria for using skills, algorithms, techniques an methods *1. Knowledge of subject specific skills and algorithms* *2. Knowledge of subject-specific techniques and methods* *3. Knowledge of criteria for determining when to use appropriate procedures*						
Metacognitive Knowledge ï knowledge of cognition in general as well as awareness and knowledge of one's own cognition *1. Strategic knowledge* *2. Cognitive tasks, including appropriate contextual and conditional knowledge* *3. Self-knowledge*						

Appendix 4.2 Table used by raters to code the performance expectations of a scorable event during the process of validation of the PET.

Table 2 Draft framework for the PET proposed for use in the analysis of performance expectations in this study.

Lower order **–** content, procedures or connections explicitly stated as required in syllabus (learned) or given in question			***Higher order*** – connections between content and/or procedures, not required by syllabus and not given in question	
ROTE i.e., no demonstration of understanding required		**COMMUNICATE UNDERSTANDING**		
A **Memorise** (i.e., knowing)	**B** **Perform-Routine-Procedures** (i.e., doing)	**C** **Demonstrate a basic understanding of memorized knowledge and routine procedures**	**D** **Analyze information/ make connections using memorized knowledge and routine procedures** (i.e., use learned or acquired knowledge) **in familiar contexts** (of syllabus or question)	**E** **Apply concepts/ make connections (**i.e. make extended use of learned or acquired information**) in new contexts** (i.e., outside of syllabus or question)
1. Recite basic science facts 2. Recall/ recognize science terms and definitions 3. Recall/ recognize scientific formulae	4. Make measurements 5. Read values from graphs 6. Compute 7. Make a scientific drawing 8. Make observations/ describe objects, processes, results 9. Use/ assemble/ handle appropriate tools, apparatus 10. Conduct routine procedures/experiments 11. Test the effects of different variables in routine experiments 12. Collect and record data 13. Organize and display data in tables, graphs or charts as instructed	14. Explain learned concepts 15. Observe and explain teacher/ student/ given demonstrations 16. Explain learned procedures and methods of science and inquiry	17. Classify and compare data (similarities and differences) 18. Organize and display data in tables, graphs or charts of own design 19. Analyze data, recognize patterns/ trends 20. Reason inductively or deductively 21. Draw conclusions 22. Identify faulty arguments or mis-representations of data 23. Generate questions or make predictions from experimental or unlearned data 24. Present analyzed information or results	25. Use and integrate science concepts 26. Test the effects of different variables 27. Recognize experimental design errors and appropriate use of controls 28. Synthesize content and ideas from several sources 29. Plan and design an investigation or experiment to address a given or generated problem or question 30. Apply and adapt science information to real-world situations 31. Apply science outside the context of science 32. Build or revise a theory or plan 33. Present applied concepts and connections

Appendix 4.3 The key used by raters to code the performance expectations of a scorable event during validation of the PET.



Note:
a Syllabus refers to core syllabus (JMB, 1984a, 1984b) and guidelines documents (DoE, 2001).

Figure 1 Key to classifying a scorable event according to the PET proposed for use in this study. Shaded boxes describe distinguishing features of the group(s) that follow and do not require decisions to be made.

Appendix 4.4 Instructions for raters during validation of the PET.

Instructions to raters for coding the performance expectations of a question paper.

<u>You need:</u>

- table of topics (Table 1)
- table of possible performance expectations - Performance Expectations Taxonomy (PET) (Table 2)
- key to arriving at categories of performance expectations (Figure 1)
- examples of the classifications of questions using PET (Figure 2)
- copy of the Biology core syllabus and guideline documents
- data capture sheet for the particular examination (Coding Sheet 1 or Coding Sheet 2)
- examination question paper
- corresponding marking memorandum

1. Each question is divided into *scorable events* (the smallest sub-question that cannot be broken down into further sub-questions).

   For example, a question that reads " 1.2.1. Identify part **B** and explain its functions. (3 marks)" is two scorable events:

   1. identify part B
   2. explain its function

2. Each scorable event needs to be classified by:

   - topic (Table 1)
   - performance expectation

     Use Figure 1 to help you to get you to the major categories (A to E) of the taxonomy.

     Table 2 may help you to describe the specific performance expectation. You do NOT have to use one of the descriptions given in this table.

3. Record topic and the performance expectation category code (A to E) and a description of the performance expectation on the data capture sheet. If you feel that a scorable event could be classified in more than one topic please note the alternates in the comments column.

4. The marks for these scorable events are proportionally allocated by looking at the mark allocation in the accompanying marking memorandum.

   For the example given in 1 above, if the marking memorandum question reads:

   " 1.2.1 pinna (1) is funnel shaped (1) to direct sound waves to the typanum (1)"

5. This question will be coded on data capture sheet (Appendix 2 or Appendix 3) as follows:

Scorable event no.	**Excel code**	**Marks**	**Topic code**	**Performance expectation**		**Comments**
				Code	**Description**	
1.2.1 i	*CA*	*1*	*54*	*A*	*Recall name of structure*	
1.2.1 ii	*CB*	*2*	*54*	*C*	*Explain the function of structure*	

Thank you!

Appendix 4.5 Table used by raters to code the topics of a scorable event during validation of the PET.

Table 1 A list of topics examined in the Biology SC Examinations 1994 - 2007 as contained in the Core Biology Syllabus (JMB, 1984; DoE (2001).

Code [a]	**Topic**
11	Biological compounds
12	Enzymes and co-enzymes
21	Angiosperm physiology: water relations
22	Angiosperm physiology: growth & development
23	Angiosperm physiology: photosynthesis
31	Cellular respiration
41	Aspects of human anatomy and physiology: nutrition
42	Aspects of human anatomy and physiology: gaseous exchange
43	Aspects of human anatomy and physiology: excretion
44	Aspects of human anatomy and physiology: co-ordination
45	Aspects of human anatomy and physiology: circulation [b]
51	Homeostasis: human nutrition
52	Homeostasis: gaseous exchange
53	Homeostasis: excretion
54	Homeostasis: co-ordination
55	Homeostasis: thermoregulation & tissue fluid
61	Population Dynamics
71	*Amoeba* and earthworm osmoregulation [c]
81	Outside of syllabus

Note:
a Topic is code/reference in this study
b IEB only, 1994 to 2003
c Removed from examination from 2003 onwards

Appendix 4.6 Handout used to orientate raters involved in validation of the PET.

Examples of five photosynthesis and cellular respiration questions and their classifications using the Performance ExpectationsTaxonomy

1. Name TWO products of photosynthesis.

Classify as A (Memorize) because knowledge of the products of photosynthesis is required by syllabus.

2. The rate of photosynthesis at different carbon dioxide concentrations was determined for plant. The results are shown in Figure 1.

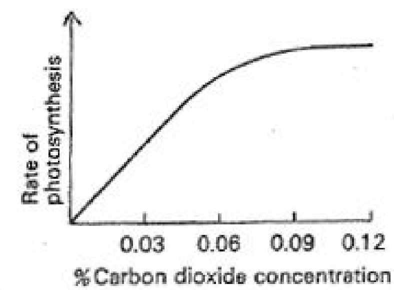


Figure 1 Photosynthesis and CO_2 concentration

What is the relationship between photosynthesis and carbon dioxide concentration?

Classify as B (Perform-routine-procedure) because the student describes data given in a graph. The student does not need to show evidence of understanding the described relationship.

3. Explain the relationship between photosynthesis and carbon dioxide concentration.

Classify as C (Demonstrate understanding of memorized knowledge and routine procedures) because the student answers by explaining the relationship (with or without the benefit of the graph in 2 above).

4. Give TWO differences between photosynthesis and cellular respiration.

Classify as D (Analyze information/make connections using memorized knowledge and routine procedures) because the student has learned about photosynthesis and respiration (i.e., familiar context), as two separate topics, but the syllabus does not require the student to make comparisons of the two processes.

Figure 2 continued on next page

Appendix 4.6 continued

5. T.W. Engelman in 1883 performed a series of experiments to investigate the role of chlorophyll in photosynthesis. He used bacteria which migrated towards oxygen as an indication that photosynthesis was taking place.

   Some of Engelman's results are shown in Figure 2 below.

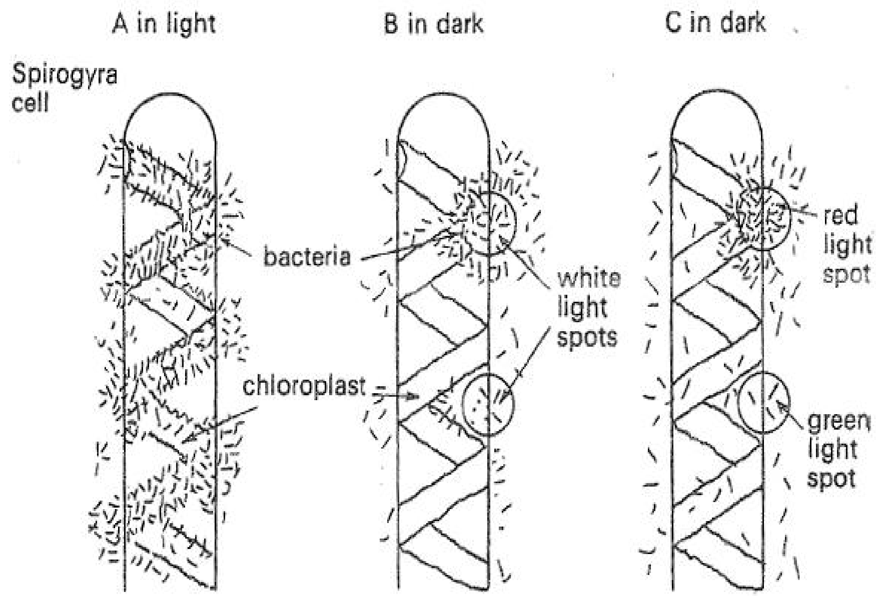


Figure 2 Engelman's experiments on the part played by chlorophyll in photosynthesis

How would the migration of these bacteria demonstrate photosynthesis?

What does the congregation of bacteria around the chloroplast in A show?

B showed that bacteria congregated around a spot of white light shining through to the chloroplast, but not around a spot shining through the cytoplasm. One hypothesis to explain this could be simply that the bacteria are attracted to green light (i.e., white light shining through a chloroplast) but not to white light. How does C refute this hypothesis?

Classify all three questions as E (Apply concepts/make connections in new contexts) because the student has to: use what he/she knows about photosynthesis and cellular respiration, make connections between these processes, and use these connections and knowledge in an entirely new context (Engelman's experiments are not required by the syllabus).

Figure 2 examples of the classifications of questions using the Performance Expectations Taxonomy proposed for use in this study.

Appendix 4.7 Data capture sheets for the analysis of 2005 Biology Higher Grade Paper 1 − validation of PET. Shaded cells were negotiated via email. (* = invalid question)

2005 BIOLOGY HIGHER GRADE Paper 1 ANALYSIS

Name: VALIDATION Date: 3rd March 2008

Scorable event no.	Excel code	Marks	Topic code	Performance expectation/rater					Performance expectation negotiated
				1	2	3	4	5	
1.1.1	C	2	41	A	A	A	A	A	A
1.1.2	D	2	41	C	B	D	D	A	D
1.1.3	E	2	42	A	B	A	D	B	A
1.1.4	F	2	23/31	D	D	D	D	C	D
1.1.5	G	2	23	A	A	A	E	C	A
1.1.6	H	2	61	C	C	A	A	A	A
1.1.7	I	2	61	A	A	A	A	A	A
1.2.1	J	1	41	A	A	A	A	A	A
1.2.2	K	1	11/41	A	A	A	A	A	A
1.2.3	L	1	23	A	A	A	A	A	A
1.2.4	M	1	41	A	A	A	A	A	A
1.2.5	N	1	61	A	A	A	A	A	A
1.2.6	O	1	61	A	A	A	A	A	A
1.3.1	P	2	41	A	A	D	A	C	A
1.3.2	Q	2	23	A	A	D	A	C	A
1.3.3	R	2	42	A	A	D	A	C	A
1.3.4	S	2	23/61	A	A	D	A	C	A
1.3.5	T	2	61	A	A	D	A	C	A
1.4.1	U	3	11/41	E	D	D	A	D	E
1.4.2	V	1	11/41	B	D	D	B	D	B
1.4.3	W	3	41	A	D	D	A	C	A
1.4.4	X	3	11/41	E	E	E	E	E	E
1.5.1	Y	2	23	A	D	B	C	B	D
1.5.2	Z	1	23	C	C	B	C	C	C
1.5.3	AA	1	11/23	A	A	B	A	A	A
1.5.4	AB	4	23	A/B	C/B	B	B	B	A
1.5.5	AC	2	23	C	D	C	C	C	C
1.6.1	AD	2	42	A	A	A	A	A	A
1.6.2	AE	3	42	A	A	A	A	A	A
1.6.3	AF	3	42	A	D	A	A	A	A
1.6.4	AG	2	42	D	D	C	E	C	D
2.1.1	AH	2	41	A	A	A	A	A	A
2.1.2	AI	1	41	A	C/D	A	A	B	A
2.1.3	AJ	2	41	C	C	C	D	C	D
2.1.4	AK	2	41	A	A	A	A	A	A
2.1.5	AL	3	41	A	A	C	A	B	A

2005 Biology Higher Grade Paper 1 analysis continued on next page

Appendix4.7 continued

Scorable event no.	Excel code	Marks	Topic code	Performance expectation/rater					Performance expectation negotiated
				1	2	3	4	5	
2.1.6	AM	2	41	A	A	D	A	B	A
2.1.7	AN	2	41	A	A	C	A	B	A
2.1.8	AO	2	41	A	A	A	A	A	A
2.1.9	AP	4	41	C	D	C	C	C	C
2.2.1	AQ	3	11/41	E	E	D	D	D	E
2.2.2	AR	2	11/41	E	E	E	E	E	E
2.2.3	AS	1	11/41	B	E	C	E	D	B
2.2.4*	AT	3	11/41	B	B	B	B	B	Invalid question
2.2.5	AU	4	11/41	D	E	D	C	C	D
2.2.6	AV	2	11	A	A	A	A	A	A
3.1.1	AW	2	11/41	E	E	D	D	B	E
3.1.2	AX	1	11/41	C	C	D	A	E	D
3.1.3	AY	4	11	A	A	A	A	A	A
3.1.4	AZ	2	11	A	A	A	A	A	A
3.2.1	BA	2	12	E	D	B	C	C	E
3.2.2	BB	1	12	E	B	B	A	C	A
3.2.3	BC	2	12	B	B	B	B	A	B
3.2.4	BD	2	12	E	D	E	A	C	E
3.2.5	BE	2	12	E	E	D	C	C	E
3.2.6	BF	2	12	E	A/B	B	C	B	E
3.2.7	BG	4	12	E	E	E	E	E	E
3.3.1	BH	2	23	B	B	B	B	B	B
3.3.2	BI	3	23	E	D/E	D	D	C	E
3.3.3	BJ	6	23	B	C	B	B	C	B
4.1.1	BK	2	61	E	C	A	A	B	A
4.1.2	BL	2	61	E	C	B	A	B	B
4.1.3	BM	1	61	D	B	D	D	C	D
4.1.4	BN	6	61	A	C	C	A	B	A
4.1.5	BO	4	61	E	E	A	E	C	E
4.2.1	BP	3	61	A	A	A	A	A	A
4.2.2	BQ	3	61	B	B	B	B	C	B
4.2.3	BR	4	61	E	C	C	E	D	E
4.3.1	BS	1	31	A	A/D	B	C	A	A
4.3.2	BTEO	4	31	E	E	C	D	C	E
4.3.3	BU	3	31	E	E	E	D	E	E
4.3.4	BV	2	31	C	C	C	C	A	C
5.1.1	BW	3	31	E	B	D	D	B/C	D
5.1.2	BX	2	31	B	B	D	B	B/C	B
5.1.3	BY	2	31	A	A	A	A	A	A
5.1.4	BZ	2	31	E	E	D	D	D	D
5.1.5	CA	2	31	A	A	A	C	C	A
5.1.6	CB	2	31	D	E	D	D	B	D
5.1.7	CC	4	31	C	D	C	D	C	C
5.2	CD	15	31/42	A	C	A	C	C	A
	CE	3	31/42	C	D	B	C	C	C

Appendix 4.8 Data capture sheets for the analysis of 2005 Biology Higher Grade Paper 2 − validation of PET. Shaded cells were negotiated via email. (* = invalid question)

2005 BIOLOGY HIGHER GRADE Paper 2 ANALYSIS

Name: VALIDATION Date: 3rd March 2008

Scorable event no.	Excel code	Marks	Topic code	Performance expectation/rater					Performance expectation negotiated
				1	2	3	4	5	
1.1.1	C	2	43	A	A	A	A	A	A
1.1.2	D	2	44	D	C	D	A	C	A
1.1.3	E	2	55	A	C	D	A	A	A
1.1.4	F	2	21	A	A	B	C	B	A
1.1.5	G	2	21	A	A	A	A	A	A
1.1.6	H	2	21	D	A	D	A	C	D
1.1.7	I	2	43	E	D	E	E	E	E
1.2.1	J	1	54	A	A	A	A	A	A
1.2.2	K	1	21	A	A	A	A	A	A
1.2.3	L	1	21	A	A	A	A	A	A
1.2.4	M	1	21	A	A	A	A	A	A
1.2.5	N	1	43	A	A	A	A	A	A
1.2.6	O	1	55	A	A	A	A	A	A
1.3.1	P	2	43	A	A	D	A	C	A
1.3.2	Q	2	43	A	A	D	A	C	A
1.3.3	R	2	21	A	A	D	A	C	A
1.3.4	S	2	21	A	A	D	A	C	A
1.3.5	T	2	54	A	A	D	A	C	A
1.4.1	U	2	21	E	D	C	C	C/D	E
1.4.2	V	1	21	A	C	B	C	C	A
1.4.3	W	6	21	E	D	C	E	B	E
1.4.4	X	1	21	E	D	D	E	A	E
1.4.5	Y	2	21	E	C	C	C	B	E
1.5.1	Z	2	44	A	A	A	A	B	A
1.5.2	AA	2	44	A	C	D	D	C	A
1.5.3	AB	2	44	C	C	C	C	C	C
1.5.4	AC	2	44	E	C	C	D	C	E
1.5.5	AD	2	44	C	C	A	E	B	C
1.6.1	AE	1	23	E	E	E	E	C	E
1.6.2	AF	2	23	E	C	E	D	C	E
1.6.3	AG	2	23	E	E	D	D	C	D
1.6.4	AH	1	23	A	A	A	A	A	A
1.6.5	AI	2	23	A	A	C	C	A	A
2.1.1	AJ	1	21	A	A	A	A	A	A
2.1.2	AK	3	21	C	A	C	C	A	C
2.1.3	AL	4	21	E	D	B	E	B	E
2.1.4	AM	4	21	C	E	C	D	C	C
2.2.1	AN	2	21	B	B	B	B	B	B
2.2.2	AO	2	21	B	B	D	B	B	C

2005 Biology Higher Grade Paper 2 analysis continued on next page

Appendix 4.8 continued

Scorable event no.	Excel code	Marks	Topic code	Performance expectation/rater					Performance expectation negotiated
				1	2	3	4	5	
2.2.3	AP	2	21	E	C	D	C	B	E
2.2.4	AQ	11	21	B	B	B	B	C	B
2.3.1	AR	1	21	D	D	B	D	C	D
2.3.2	AS	2	21	D	C	C	C	B	D
2.3.3	AT	3	21	D	A	B	C	C	D
3.1.1	AU	8	43	B	B	B	B	A	B
3.1.2	AV	3	43	A	A	A	A	A	A
3.2.1	AW	1	43	A	A	A	A	A	A
3.2.2	AX	3	43	A	A	A	A	A	A
3.2.3	AY	1	43	A	A	A	A	C	A
3.2.4	AZ	6	43	C	C	C	C	B	C
3.2.5	BA	3	43	D	B	D	D	B	D
3.2.6	BB	1	43	A	C	C	A	B	A
3.2.7	BC	4	43	C	D	C	D	C	C
3.2.8*	BD	5	43	E	E	E	E	E	Invalid question
4.1.1	BE	4	55	D	D	C	D	E	D
4.1.2	BF	3	55	D	C	D	D	C	D
4.1.3	BG	3	55	D	C	A	C/E	C	C
4.2.1	BH	2	55	A	A	A	A	A	A
4.2.2	BI	1	55	A	C	C	A	B	A
4.2.3	BJ	4	55	C	C	C	C	C	C
4.2.4 i	BK	1	44/55	D	A	C	A	A	D
4.2.4 ii	BL	1	44/55	D	A	A	A	A	D
4.2.4 iii	BM	5	44/55	D	C	C	C	B	D
4.2.4 iv	BN	1	44	D	D	D	E	B	D
4.2.4 v	BO	2	44	D	C	C	C	C	D
4.3.1	BP	2	44	D	A	D	A	C	A
4.3.2	BQ	3	44	E	B	E	A	D	E
4.3.3	BR	3	44	E	B	C	E	C	A
5.1.1	BS	2	44	A	A	A	A	A	A
5.1.2 i	BTEO	3	44	E	E	D	C	E	E
5.1.2 ii	BU	2	44	E	C	C	D	E	E
5.1.3 i	BV	1	44	D	A	A	A	C	D
5.1.3 ii	BW	2	44	D	C	C	C	C	D
5.1.4	BX	2	44	E	C	C	D	E	E
5.1.5	BY	2	44	E	E	E	E	E	E
5.1.6 i	BZ	1	44	E	C	E	A	E	E
5.1.6 ii	CA	2	44	E	C	E	C	E	E
5.2	CB	15	44	A	C	A	E	E	A
	CC	3	44	C	B	B	E	E	C

Appendix 4.9 Table used by raters to check classifications during the PET post-validation. Shaded cells were contested and negotiated. (* = invalid question).

2005 BIOLOGY HIGHER GRADE Paper 2 CROWE ANALYSIS for comment

Name: POST-VALIDATION Date: 4th March 2008

Scorable event no.	Excel code	Marks	Topic code	Performance expectation code	Explanation
1.1.1	C	2	43	A	Recall of fact in the CBS
1.1.2	D	2	44	A	Recall descriptions of terms in the CBS
1.1.3	E	2	55	A	Recall of facts in the CBS
1.1.4	F	2	21	A	Recall of fact in the CBS
1.1.5	G	2	21	A	Recall of fact in the CBS
1.1.6	H	2	21	D	Recall facts from different sections of work (familiar) and put facts together (connect) to decide on the functions of thorns
1.1.7	I	2	43	E	Must understand what conditions influence the amount of urine produced and use this knowledge (apply) to match the concentration of urine and conditions given (unfamiliar)
1.2.1	J	1	54	A	Recall of term in the CBS
1.2.2	K	1	21	A	Recall of term in the CBS
1.2.3	L	1	21	A	Recall of term in the CBS
1.2.4	M	1	21	A	Recall of term in the CBS
1.2.5	N	1	43	A	Recall of term in the CBS
1.2.6	O	1	55	A	Recall of term in the CBS
1.3.1	P	2	43	A	Recall of terms/concepts. Not a D because the matching is in how the answer is presented not in the answer itself.
1.3.2	Q	2	43	A	Recall of terms/concepts. Not a D because the matching is in how the answer is presented not in the answer itself.
1.3.3	R	2	21	A	Recall of terms/concepts. Not a D because the matching is in how the answer is presented not in the answer itself.
1.3.4	S	2	21	A	Recall of terms/concepts. Not a D because the matching is in how the answer is presented not in the answer itself.

2005 Biology Higher Grade Paper 2 continued on next page

Appendix 4.9 continued

Scorable event no.	Excel code	Marks	Topic code	Performance expectation code	Explanation
1.3.5	T	2	54	A	Recall of terms/concepts. Not a D because the matching is in how the answer is presented not in the answer itself.
1.4.1	U	2	21	E	Experiment is not prescribed by syllabus (unfamiliar)
1.4.2	V	1	21	A	Recall of fact in the CBS.
1.4.3	W	6	21	E	Experiment is not prescribed by syllabus (unfamiliar)
1.4.4	X	1	21	E	Experiment is not prescribed by syllabus (unfamiliar)
1.4.5	Y	2	21	E	Experiment is not prescribed by syllabus (unfamiliar)
1.5.1	Z	2	44	A	Recall fact in the CBS.
1.5.2	AA	2	44	A	Recall/recognize facts in the CBS
1.5.3	AB	2	44	C	Show understanding of relationship between light intensity and pupil size – in the CBS.
1.5.4	AC	2	44	E	Graph showing change in lens shape over a period of time is unfamiliar. Students would only have learnt about accommodation as more and less convex lens and to answer this would need to extrapolate the in-between lens shapes over a period of time.
1.5.5	AD	2	44	C	Show understanding of a memorized fact in the CBS
1.6.1	AE	1	23	E	The effect of differing wavelengths of light, other than green, are not required so this question is unfamiliar.
1.6.2	AF	2	23	E	The effect of differing wavelengths of light, other than green, are not required so this question is unfamiliar.
1.6.3	AG	2	23	D	Connection made between key and graph which are given in question (familiar). This is not unfamiliar (E) because all that is required is to read effectiveness and indigo light which are on graph.

2005 Biology Higher Grade Paper 2 continued on next page

Appendix 4.9 continued

Scorable event no.	**Excel code**	**Marks**	**Topic code**	**Performance expectation code**	**Explanation**
1.6.4	AH	1	23	A	Recall of fact in the CBS.
1.6.5	AI	2	23	A	Recall of fact in the CBS.
2.1.1	AJ	1	21	A	Recall of fact in the CBS.
2.1.2	AK	3	21	C	Recall of fact AND explanation of fact in the CBS.
2.1.3	AL	4	21	E	Not classified as C because the syllabus does not ask for reliable functioning of this particular experiment – it is a general skill. The context is therefore treated as unfamiliar.
2.1.4	AM	4	21	C	Explanation of effect of photosynthesis in guard cells on size of stomatal pore and therefore rate of transpiration – in the CBS (familiar)
2.2.1	AN	2	21	B	Perform a calculation (procedure)
2.2.2	AO	2	21	C	Read a value from a table (procedure) AND link it to relationship between transpiration and environmental conditions – in the CBS.
2.2.3	AP	2	21	E	Explanation is required for results obtained in an unfamiliar context.
2.2.4	AQ	11	21	B	Draw a graph (procedure)
2.3.1	AR	1	21	D	Interpret the different shapes of the cells with respect to whether they have gained or lost water and connect this to why they gained or lost water (the concentration of the solutions). I did not make this an E because the question shows the three classic states plant cells can be in (therefore familiar) depending on the concentration of the solution in which they occur – requirement of syllabus.
2.3.2	AS	2	21	D	As above
2.3.3	AT	3	21	D	As above
3.1.1	AU	8	43	B	Draw a kidney (procedure) the structure of which is in the CBS.
3.1.2	AV	3	43	A	Recall of facts in the CBS.

2005 Biology Higher Grade Paper 2 continued on next page

Appendix 4.9 continued

Scorable event no.	**Excel code**	**Marks**	**Topic code**	**Performance expectation code**	**Explanation**
3.2.1	AW	1	43	A	Recognition and recall the name of a nephron – in the CBS
3.2.2	AX	3	43	A	Recall/recognition of parts of a nephron – in the CBS.
3.2.3	AY	1	43	A	Recognize part of a nephron and recall its function.
3.2.4	AZ	6	43	C	Explain the adaptation of part of nephron – in the CBS.
3.2.5	BA	3	43	D	Perform a calculation (procedure) using data in question (familiar) but in order to this must understand the flow of liquid through a nephron (familiar).
3.2.6	BB	1	43	A	Fact is in the CBS. This requires no knowledge and student could guess.
3.2.7	BC	4	43	C	Explain facts about the functioning of the nephron in the CBS.
3.2.8*	BD	5	43	E	Question is invalid - all students credited for this question.
4.1.1	BE	4	55	D	Relationships between thermo-regulation, extremeties and shape of an organism – in the CBS. Even though students might not have been exposed specifically to the seal the diagram given in the question clarifies its extremities and its shape (familiar)
4.1.2	BF	3	55	D	Knowledge of heat-exchangers in the CBS needs to be related to diagram given (familiar).
4.1.3	BG	3	55	C	Explanation of the function of heat exchanger. Initially I had this as a D but it does not require analysis or putting together of different bits of information.
4.2.1	BH	2	55	A	Recall of names of parts of skin – in the CBS.
4.2.2	BI	1	55	A	Recall functioning of part of skin – in the CBS.

2005 Biology Higher Grade Paper 2 continued on next page

Appendix 4.9 continued

Scorable event no.	**Excel code**	**Marks**	**Topic code**	**Performance expectation code**	**Explanation**
4.2.3	BJ	4	55	C	Explain the reason for particular functioning of part of skin – part of syllabus.
4.2.4 i	BK	1	44/55	D	Changes in skin with changing temperature and structure of the skin AND the effect of hormones on skin structures are required in different sections of the syllabus. Need to make the connection between the two sections.
4.2.4 ii	BL	1	44/55	D	See above
4.2.4 iii	BM	5	44/55	D	See above
4.2.4 iv	BN	1	44	D	Inter-relationship between adrenalin and thyroxine not in the CBS but student uses knowledge of adrenalin AND knowledge of thyroxine (familiar)
4.2.4 v	BO	2	44	D	See above
4.3.1	BP	2	44	A	Answer given in text and recall of fact in the CBS.
4.3.2	BQ	3	44	E	Explanations of conditions not in the CBS (unfamiliar)
4.3.3	BR	3	44	A	Answer given in question.
5.1.1	BS	2	44	A	Recognition and recall of structure in the CBS.
5.1.2 i	BTEO	3	44	E	The context of this question is unfamiliar.
5.1.2 ii	BU	2	44	E	See above.
5.1.3 i	BV	1	44	D	The structure and functioning of the brain and the sense organs are treated in the same section but independently by the syllabus (familiar). Question is answered by connecting the information.
5.1.3 ii	BW	2	44	D	Explanation of the connection described above.
5.1.4	BX	2	44	E	Functions of the medulla oblongata are in the CBS but need to connect this knowledge to knowledge of what will cause death in an organism (unfamiliar)

2005 Biology Higher Grade Paper 2 continued on next page

Appendix 4.9 continued

Scorable event no.	Excel code	Marks	Topic code	Performance expectation code	Explanation
5.1.5	BY	2	44	E	Distribution of sense organs not in the CBS (unfamiliar)
5.1.6 i	BZ	1	44	E	See above
5.1.6 ii	CA	2	44	E	See above
5.2	CB	15	44	A	Facts in the CBS. I would have classified this as a C BUT these facts are marked correct even if they are not arranged to make the best sense. The arrangement is credited only in the three marks below.
	CC	3	44	C	This refers to the way in which recalled facts are put together. If students understand the facts they will arrange them in an "organized" way and will score higher than students who have little understanding of the facts and so cannot arrange them in an organized way. It cannot be anything higher than a C because students were required to learn this work by the syllabus

Appendix 5.1 List of SC Biology examination question papers analyzed in this study.

Year	Examining body	Grade	Source of paper
1994	Transvaal	HG	Scheltema and Myburgh (1995)
1994	Cape	HG	Scheltema and Myburgh (1995)
1994	Natal	HG	Scheltema and Myburgh (1995)
1994	Orange Free State	HG	Scheltema and Myburgh (1995)
1994	HOD	HG	Scheltema and Myburgh (1995)
1994	HOR	HG	Scheltema and Myburgh (1995)
1994	DET	HG	Scheltema and Myburgh (1995)
1994	HOA	HG	Scheltema and Myburgh (1995)
1994	IEB	HG	Scheltema and Myburgh (1995)
1995	Transvaal	HG	Scheltema and Myburgh (1996)
1995	Cape	HG	Scheltema and Myburgh (1996)
1995	Natal	HG	Scheltema and Myburgh (1996)
1995	Orange Free State	HG	Scheltema and Myburgh (1996)
1995	HOD	HG	Scheltema and Myburgh (1996)
1995	HOR	HG	Scheltema and Myburgh (1996)
1995	DET	HG	Scheltema and Myburgh (1996)
1995	HOA	HG	Scheltema and Myburgh (1996)
1995	IEB	HG	Scheltema and Myburgh (1996)
1996	Gauteng	HG	Scheltema and Myburgh (1997)
1996	Western Cape	HG	Scheltema and Myburgh (1997)
1996	Eastern Cape	HG	Scheltema and Myburgh (1997)
1996	Northern Cape	HG	Scheltema and Myburgh (1997)
1996	KwaZulu-Natal	HG	Scheltema and Myburgh (1997)
1996	Free State	HG	Scheltema and Myburgh (1997)
1996	Mpumalanga	HG	Scheltema and Myburgh (1997)
1996	Northern Province	HG	Scheltema and Myburgh (1997)
1996	North West	HG	Scheltema and Myburgh (1997)
1996	IEB	HG	Scheltema and Myburgh (1997)
1997	Gauteng	HG	Scheltema and Myburgh (1998)
1997	Western Cape	HG	Scheltema and Myburgh (1998)
1997	Eastern Cape	HG	Scheltema and Myburgh (1998)
1997	Northern Cape	HG	Scheltema and Myburgh (1998)
1997	KwaZulu-Natal	HG	Scheltema and Myburgh (1998)
1997	Free State	HG	Scheltema and Myburgh (1998)
1997	Mpumalanga	HG	Scheltema and Myburgh (1998)
1997	Northern Province	HG	Scheltema and Myburgh (1998)
1997	North West	HG	Scheltema and Myburgh (1998)
1997	IEB	HG	Scheltema and Myburgh (1998)
1998	Gauteng	HG	Scheltema and Myburgh (1999)
1998	Western Cape	HG	Scheltema and Myburgh (1999)
1998	Eastern Cape	HG	Scheltema and Myburgh (1999)
1998	Northern Cape	HG	Scheltema and Myburgh (1999)
1998	KwaZulu-Natal	HG	Scheltema and Myburgh (1999)
1998	Free State	HG	Scheltema and Myburgh (1999)

Appendix 5.1 continued

1998	Mpumalanga	HG	Scheltema and Myburgh (1999)
1998	Northern Province	HG	Scheltema and Myburgh (1999)
1998	North West	HG	Scheltema and Myburgh (1999)
1998	IEB	HG	Scheltema and Myburgh (1999)
1999	Gauteng	HG	Scheltema and Myburgh (2000)
1999	Western Cape	HG	Scheltema and Myburgh (2000)
1999	Eastern Cape	HG	Scheltema and Myburgh (2000)
1999	Northern Cape	HG	Scheltema and Myburgh (2000)
1999	KwaZulu-Natal	HG	Scheltema and Myburgh (2000)
1999	Free State	HG	Scheltema and Myburgh (2000)
1999	Mpumalanga	HG	Scheltema and Myburgh (2000)
1999	Northern Province	HG	Scheltema and Myburgh (2000)
1999	North West	HG	Scheltema and Myburgh (2000)
1999	IEB	HG	Scheltema and Myburgh (2000)
2000	Gauteng	HG	Scheltema and Myburgh (2001)
2000	Western Cape	HG	Scheltema and Myburgh (2001)
2000	Eastern Cape	HG	Scheltema and Myburgh (2001)
2000	Northern Cape	HG	Scheltema and Myburgh (2001)
2000	KwaZulu-Natal Paper 1	HG	Scheltema and Myburgh (2001)
2000	KwaZulu-Natal Paper 2	HG	Scheltema and Myburgh (2001)
2000	Free State	HG	Scheltema and Myburgh (2001)
2000	Mpumalanga	HG	Scheltema and Myburgh (2001)
2000	Northern Province	HG	Scheltema and Myburgh (2001)
2000	North West	HG	Scheltema and Myburgh (2001)
2000	IEB	HG	Scheltema and Myburgh (2001)
2001	National DoE Paper 1	HG & SG	DoE (2001c,2001e)
2001	National DoE Paper 2	HG & SG	DoE (2001d, 2001f)
2001	IEB	HG & SG	IEB (2001a, 2001b)
2002	National DoE Paper 1	HG & SG	DoE (2002d,2002f)
2002	National DoE Paper 2	HG & SG	DoE (2002e, 2002g)
2002	IEB	HG & SG	IEB (2002a, 2002b)
2003	National DoE Paper 1	HG & SG	DoE (2003b,2003d)
2003	National DoE Paper 2	HG & SG	DoE (2003c, 2003e)
2003	IEB	HG & SG	IEB (2003a, 2003b)
2004	National DoE Paper 1	HG & SG	DoE (2004e,2004g)
2004	National DoE Paper 2	HG & SG	DoE (2004f, 2004h)
2004	IEB	HG & SG	IEB (2004a, 2004b)
2005	National DoE Paper 1	HG & SG	DoE (2005b,2005d)
2005	National DoE Paper 2	HG & SG	DoE (2005c, 2005e)
2005	IEB	HG & SG	IEB (2005a, 2005b)
2006	National DoE Paper 1	HG & SG	DoE (2006b,2006d)
2006	National DoE Paper 2	HG & SG	DoE (2006c, 2006e)
2006	IEB	HG & SG	IEB (2006a, 2006b)
2007	National DoE Paper 1	HG & SG	DoE (2007a,2007c)
2007	National DoE Paper 2	HG & SG	DoE (2007b, 2007d)
2007	IEB	HG & SG	IEB (2007b, 2007c)

Appendix 5.2 The coding form used to extract data from examination question papers. The terminology used for Section A question types differs slightly from that used in the final analyses.

BIOLOGY EXAMINATION QUESTION PAPER DATA CAPTURE SHEETS

Year:

Examining body:

Number of papers:

Number of this paper:

Time allowed:

Maximum marks:

Maximum marks - Section A:

Question types:

MCQ	Matching columns/diagrams	One term	True/False

Section B:

Section C:

No. of possible answers in MCQs:

Section with choice:

Notes on source(s) of challenge:

Total pages	Date captured	Date coded

Page 1 of

Appendix 5.2 continued

Scorable event number	Mark allocated	Question type		Question characteristics					Answer characteristics (free response)				COMB		
		Choose correct answer	Free response	One sentence	Two-three sentences	> three sentences	Use of non text	Source of challenge	One/two terms	Short	Extended	Non-text	Topic	Performance expectation	

Page of

Appendix 5.3 The number of scorable events recognized in the biology question papers analyzed in this study.

A. 1994 to 2000 Biology HG

All examinations consisted of one paper three hours long, except the 2000 KwaZulu-Natal examination which was two papers each two hours long.

Examination body	Year						
	1994	1995	1996	1997	1998	1999	2000
Transvaal [a]	88	110					
Cape [a]	104	127					
Natal [a]	93	78					
Orange Free State [a]	117	120					
HOD [b]	96	93					
HOR [b]	109	106					
DET [b]	134	128					
National Examination Board [c]	97	97					
Gauteng [d]			101	107	134	108	133
Western Cape [d]			121	109	121	128	130
Eastern Cape [d]			132	111	139	142	138
Northern Cape [d]			102	118	80	117	120
Kwazulu- Natal [d]			97	106	113	95	143 [e]
Free State [d]			100	108	105	99	110
Mpumalanga [d]			112	128	96	123	130
Northern Province [d]			111	132	114	141	136
Northwest Province [d]			103	101	150	134	134
IEB	89	80	72	88	85	79	86

Note:
a Provincial Education Department
b DoE and Culture
c National Examination Board
d Provincial Education Department
e Paper 1 and Paper 2 combined

Appendix 5.3 continued

B. 2001 to 2007 Biology HG and SG

The national DoE examinations consisted of two papers each two hours long and the IEB examination consisted of one paper of three hours long. Scorable events for the two DoE examinations have been combined below for comparative purposes.

Year	DoE HG	DoE SG	IEB HG	IEB SG
2001	160	122	76	71
2002	175	142	87	82
2003	148	127	96	80
2004	162	126	91	55
2005	158	134	90	75
2006	136	119	95	89
2007	155	143	92	71

Appendix 5.4 The coding form used to compile the profiles of examination question papers. The terminology used for Section A question types may differ slightly from that used in the final analyses and in the Excel spreadsheets.

BIOLOGY EXAMINATION QUESTION PAPER COMPUTER CODING SHEETS **(PAGE 1)**

					Section A								B	Section C			Structure of questions							
Year	Examining body	No. of paper	No. of profile	Total marks	MCQ	Matching columns/ diagrams	Item/Statement	Terminology/ one word	Missing words/labels	Incorrect/correct labels	True/false or appropriate/not appropriate	Short questions	Short questions	MCQ	Short questions	Essay	Select answer from choices given	Free response	Questions one sentence	Questions two to three sentences	> Three sentences	Use of diagrams	Use of graphs	Use of tables

Appendix 5.4 continued

BIOLOGY EXAMINATION QUESTION PAPER COMPUTER CODING SHEETS **(PAGE 2)**

				Structure of questions cont.			Structure of answers						Topics												
Year	Examining body	No. of paper	No. of profile	Use of diagrams	Use of graphs	Use of tables	One/two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables	Biochemistry	Photosynthesis	Respiration	Human nutrition	Human gaseous exchange	Population dynamics	Plant hormones	Plant water relationships	Human excretion	Human co-ordination	Thermoregulation	Human circulation	Out of syllabus

Appendix 5.4 continued

BIOLOGY EXAMINATION QUESTION PAPER COMPUTER CODING SHEETS **(PAGE 3)**

				Performance expectations					Content																	
Year	Examining body	No. of paper	No. profile	Memorize	Perform procedures	Explain	Analyze	Apply	COMB 1	COMB 2	COMB 3	COMB 4	COMB 5	COMB 6	COMB 7	COMB 8	COMB 9	COMB 10	COMB 11	COMB 12	COMB 13	COMB 14	COMB 15	COMB 16	COMB 17	COMB 18

Appendix 5.4 continued

BIOLOGY EXAMINATION QUESTION PAPER DATA CODING SHEETS **(PAGE 4)**

				Content cont.																						
Year	Examining body	No. of paper	No. of profile	COMB 19	COMB 20	COMB 21	COMB 22	COMB 23	COMB 24	COMB 25	COMB 26	COMB 27	COMB 28	COMB 29	COMB 30	COMB 31	COMB 32	COMB 33	COMB 34	COMB 35	COMB 36	COMB 37	COMB 38	COMB 39	COMB 40	COMB 41

Appendix 5.4 continued

BIOLOGY EXAMINATION QUESTION PAPER DATA CODING SHEETS **(PAGE 5)**

				Content cont.																						
Year	Examining body	No. of paper	No. of profile	COMB 42	COMB 43	COMB 44	COMB 45	COMB 6	COMB 47	COMB 48	COMB 49	COMB 50	COMB 51	COMB 52	COMB 53	COMB 54	COMB 55	COMB 56	COMB 57	COMB 58	COMB 59	COMB 60	COMB 61	COMB 62	COMB 63	COMB 64

Appendix 5.4 continued

BIOLOGY EXAMINATION QUESTION PAPER DATA CODING SHEETS (PAGE 6)

				Content		
Year	Examining body	No. of papers	No. of profile	COMB 65	BOK	DOK

Appendix 5.5 Summaries of the questions used to compile the profile(s) for each of the SC Biology question papers analyzed in this study.

A. 1994 HG

Examining Body	Profile	Questions
Transvaal	1	1; 2; 3; 4; 5
Transvaal	2	1; 2; 3; 4; 6
Cape	1	1; 2; 3; 4; 5
Cape	2	1; 2; 3; 4; 6
Natal	1	1; 2; 3; 4; 5
Natal	2	1; 2; 3; 4; 6
Orange Free State	1	1; 2; 3; 4; 5
Orange Free State	2	1; 2; 3; 4; 6
HOD	1	1; 2; 3; 4; 5
HOD	2	1; 2; 3; 4; 6
HOR	1	1; 2; 3; 4; 5
HOR	2	1; 2; 3; 4; 6
DET	1	1; 2; 3; 4; 5
HOA	1	1; 2; 3; 4; 5
HOA	2	1; 2; 3; 4; 6
IEB	1	1; 2; 3; 4; 5
IEB	2	1; 2; 3; 4; 6

Appendix 5.5 continued

B. 1995 HG

Examining Body	Profile	Questions
Transvaal	1	1; 2; 3; 4; 5
Transvaal	2	1; 2; 3; 4; 6
Cape	1	1; 2; 3; 4; 5
Cape	2	1; 2; 3; 4; 6
Natal	1	1; 2; 3; 4; 5
Natal	2	1; 2; 3; 4; 6
Orange Free State	1	1; 2; 3; 4; 5
Orange Free State	2	1; 2; 3; 4; 6
HOD	1	1; 2; 3; 4; 5
HOD	2	1; 2; 3; 4; 6
HOR	1	1; 2; 3; 4; 5
HOR	2	1; 2; 3; 4; 6
DET	1	1; 2; 3; 4; 5
DET	2	1; 2; 3; 4; 6
HOA	1	1; 2; 3; 4; 5
HOA	2	1; 2; 3; 4; 6
IEB	1	1; 2; 3; 4; 5
IEB	2	1; 2; 3; 4; 6

Appendix 5.5 continued

C. 1996 HG

Examining Body	Profile	Questions [a]
Gauteng	1	1; 2; 3; 4; 5
Western Cape	1	1; 2; 3; 4; 5
Western Cape	2	1; 2; 3; 4; 6
Eastern Cape	1	1; 2; 3; 4; 5
Eastern Cape	2	1; 2; 3; 4; 6
Northern Cape	1	1; 2; 3; 4; 5
Northern Cape	2	1; 2; 3; 4; 6
KwaZulu-Natal	1	1; 2; 3; 4; 5
KwaZulu-Natal	2	1; 2; 3; 4; 6
Free State	1	1; 2; 3; 4; 5
Free State	2	1; 2; 3; 4; 6
Mpumalanga	1	1; 2; 3; 4; 5.1; 5.2.1
Mpumalanga	2	1; 2; 3; 4; 5.1; 5.2.2
Northern Province	1	1; 2; 3; 4; 5
Northern Province	2	1; 2; 3; 4; 6
North West	1	1; 2; 3; 4; 5
North West	2	1; 2; 3; 4; 6
IEB	1	1; 2.1; 2.2; 2.3.1; 2.3.2; 3; 4.1; 4.2; 4.3; 4.4a / 4.4b [b]; 5; 6
IEB	2	1; 2.1; 2.2; 2.3.1; 2.3.2; 3; 4.1; 4.2; 4.3; 4.4a / 4.4b [b]; 5; 7
IEB	3	1; 2.1; 2.2; 2.3.3; 3; 4.1; 4.2; 4.3; 4.4a / 4.4b [b]; 5; 6
IEB	4	1; 2.1; 2.2; 2.3.3; 3; 4.1; 4.2; 4.3; 4.4a / 4.4b [b]; 5; 7

Note:

a Sub-question numbers are given only where there is a choice within a question.

b Choice of questions have the exact same profile.

Appendix 5.5 continued

D. 1997 HG

Examining Body	Profile	Questions [a]
Gauteng	1	1; 2; 3; 4; 5
Western Cape	1	1; 2; 3; 4; 5
Western Cape	2	1; 2; 3; 4; 6
Eastern Cape	1	1; 2; 3; 4; 5
Eastern Cape	2	1; 2; 3; 4; 6
Northern Cape	1	1; 2; 3; 4; 5
Northern Cape	2	1; 2; 3; 4; 6
KwaZulu-Natal	1	1; 2; 3; 4; 5
KwaZulu-Natal	2	1; 2; 3; 4; 6
Free State	1	1; 2; 3; 4; 5
Free State	2	1; 2; 3; 4; 6
Mpumalanga	1	1; 2; 3; 4; 5.1; 5.2.1
Mpumalanga	2	1; 2; 3; 4; 5.1; 5.2.2
Northern Province	1	1; 2; 3; 4; 5
Northern Province	2	1; 2; 3; 4; 6
North West	1	1; 2; 3; 4; 5
North West	2	1; 2; 3; 4; 6
IEB	1	1; 2.1; 2.2; 2.3; 3; 4; 5.1; 5.2; 5.3; 6
IEB	2	1; 2.1; 2.2; 2.3; 3; 4; 5.1; 5.2; 5.3; 7
IEB	3	1; 2.1; 2.2; 2.3; 3; 4; 5.1; 5.2; 5.4; 6
IEB	4	1; 2.1; 2.2; 2.3; 3; 4; 5.1; 5.2; 5.4; 7
IEB	5	1; 2.1; 2.2; 2.4; 3; 4; 5.1; 5.2; 5.3; 6
IEB	6	1; 2.1; 2.2; 2.4; 3; 4; 5.1; 5.2; 5.3; 7
IEB	7	1; 2.1; 2.2; 2.4; 3; 4; 5.1; 5.2; 5.4; 6
IEB	8	1; 2.1; 2.2; 2.4; 3; 4; 5.1; 5.2; 5.4; 7

Note:

a Sub-question numbers are given only where there is a choice within a question.

Appendix 5.5 continued

E. 1998 HG

Examining Body	**Profile**	**Questions** [a]
Gauteng	1	1; 2; 3; 4; 5
Western Cape	1	1; 2; 3; 4; 5
Western Cape	2	1; 2; 3; 4; 6
Eastern Cape	1	1; 2; 3; 4; 5
Eastern Cape	2	1; 2; 3; 4; 6
Northern Cape	1	1; 2; 3; 4; 5
Northern Cape	2	1; 2; 3; 4; 6
KwaZulu-Natal	1	1; 2; 3; 4; 5
KwaZulu-Natal	2	1; 2; 3; 4; 6
Free State	1	1; 2; 3; 4; 5
Free State	2	1; 2; 3; 4; 6
Mpumalanga	1	1; 2; 3; 4; 5.1; 5.2.1
Mpumalanga	2	1; 2; 3; 4; 5.1; 5.2.2
Northern Province	1	1; 2; 3; 4; 5
Northern Province	2	1; 2; 3; 4; 6
North West	1	1; 2; 3; 4; 5
North West	2	1; 2; 3; 4; 6
IEB	1	1; 2.1; 2.2; 2.3 / 2.4 [b]; 3; 4; 5.1; 5.2; 5.3; 5.4; 6
IEB	2	1; 2.1; 2.2; 2.3 / 2.4 [b]; 3; 4; 5.1; 5.2; 5.3; 5.4; 7
IEB	3	1; 2.1; 2.2; 2.3 / 2.4 [b]; 3; 4; 5.1; 5.2; 5.3; 5.4; 8
IEB	4	1; 2.1; 2.2; 2.3 / 2.4 [b]; 3; 4; 5.1; 5.2; 5.3; 5.5; 6
IEB	5	1; 2.1; 2.2; 2.3 / 2.4 [b]; 3; 4; 5.1; 5.2; 5.3; 5.5; 7
IEB	6	1; 2.1; 2.2; 2.3 / 2.4 ; 3; 4; 5.1; 5.2; 5.3; 5.5; 8

Note:
a Sub-question numbers are given only where there is a choice within a question.
b Choice of questions have the exact same profile.

Appendix 5.5 continued

F. 1999 HG

Examining Body	Profile	Questions [a]
Gauteng	1	1; 2; 3; 4; 5
Western Cape	1	1; 2; 3; 4; 5
Western Cape	2	1; 2; 3; 4; 6
Eastern Cape	1	1; 2; 3; 4; 5
Eastern Cape	2	1; 2; 3; 4; 6
Northern Cape	1	1; 2; 3; 4; 5
Northern Cape	2	1; 2; 3; 4; 6
KwaZulu-Natal	1	1; 2; 3; 4; 5
KwaZulu-Natal	2	1; 2; 3; 4; 6
Free State	1	1; 2; 3; 4; 5
Free State	2	1; 2; 3; 4; 6
Mpumalanga	1	1; 2; 3; 4; 5; 6.1
Mpumalanga	2	1; 2; 3; 4; 5; 6.2
Northern Province	1	1; 2; 3; 4; 5
Northern Province	2	1; 2; 3; 4; 6
North West	1	1; 2; 3; 4; 5
North West	2	1; 2; 3; 4; 6
IEB	1	1; 2.1; 2.2; 2.3 / 2.4; 3; 4; 5.1; 5.2; 5.3; 5.4; 5.5; 6
IEB	2	1; 2.1; 2.2; 2.3 / 2.4; 3; 4; 5.1; 5.2; 5.3; 5.4; 5.5; 7
IEB	3	1; 2.1; 2.2; 2.3 / 2.4; 3; 4; 5.1; 5.2; 5.3; 5.4; 5.6; 6
IEB	4	1; 2.1; 2.2; 2.3 / 2.4; 3; 4; 5.1; 5.2; 5.3; 5.4; 5.6; 7

Note:
a Sub-question numbers are given only where there is a choice within a question

Appendix 5.5 continued

G. 2000 HG

Examining Body	Profile	Questions [a]
Gauteng	1	1; 2; 3; 4; 5
Western Cape	1	1; 2; 3; 4; 5
Western Cape	2	1; 2; 3; 4; 6
Eastern Cape	1	1; 2; 3; 4; 5
Eastern Cape	2	1; 2; 3; 4; 6
Northern Cape	1	1; 2; 3; 4; 5
Northern Cape	2	1; 2; 3; 4; 6
KwaZulu-Natal Paper 1	1	1; 2; 3; 4; 5
KwaZulu-Natal Paper 2	1	1; 2; 3; 4; 5
KwaZulu-Natal combined [b]	1	Paper 1: 1; 2; 3; 4; 5 and Paper 2: 1; 2; 3; 4; 5
Free State	1	1; 2; 3; 4; 5
Free State	2	1; 2; 3; 4; 6
Mpumalanga	1	1; 2; 3; 4; 5; 6.1
Mpumalanga	2	1; 2; 3; 4; 5; 6.2
Northern Province	1	1; 2; 3; 4; 5
Northern Province	2	1; 2; 3; 4; 6
North West	1	1; 2; 3; 4; 5
North West	2	1; 2; 3; 4; 6
IEB	1	1; 2.1; 2.2; 2.3; 2.4; 3; 4; 5.1; 5.2; 5.3; 5.4; 6
IEB	2	1; 2.1; 2.2; 2.3; 2.4; 3; 4; 5.1; 5.2; 5.3; 5.4; 7
IEB	3	1; 2.1; 2.2; 2.3; 2.4; 3; 4; 5.1; 5.2; 5.3; 5.5; 6
IEB	4	1; 2.1; 2.2; 2.3; 2.4; 3; 4; 5.1; 5.2; 5.3; 5.5; 7
IEB	5	1; 2.1; 2.2; 2.3; 2.5; 3; 4; 5.1; 5.2; 5.3; 5.4; 6
IEB	6	1; 2.1; 2.2; 2.3; 2.5; 3; 4; 5.1; 5.2; 5.3; 5.4; 7
IEB	7	1; 2.1; 2.2; 2.3; 2.5; 3; 4; 5.1; 5.2; 5.3; 5.5; 6
IEB	8	1; 2.1; 2.2; 2.3; 2.5; 3; 4; 5.1; 5.2; 5.3; 5.5; 7

Note:

a Sub-question numbers are given only where there is a choice within a question.

b Paper 1 and Paper 2 combined for comparative purposes.

Appendix 5.5 continued

H. 2001 to 2007 National HG (a) and SG (b)

a HG

Year	Paper	Profile	Questions
2001	Paper 1	1	1; 2; 3; 4; 5
2001	Paper 2	1	1; 2; 3; 4; 5
2001	Combined [a]	1	Papers 1 & 2
2002	Paper 1	1	1; 2; 3; 4; 5
2002	Paper 2	1	1; 2; 3; 4; 5
2002	Combined	1	Papers 1 & 2
2003	Paper 1	1	1; 2; 3; 4; 5
2003	Paper 2	1	1; 2; 3; 4; 5
2003	Combined	1	Papers 1 & 2
2004	Paper 1	1	1; 2; 3; 4; 5
2004	Paper 2	1	1; 2; 3; 4; 5
2004	Combined	1	Papers 1 & 2
2005	Paper 1	1	1; 2; 3; 4; 5
2005	Paper 2	1	1; 2; 3; 4; 5
2005	Combined	1	Papers 1 & 2
2006	Paper 1	1	1; 2; 3; 4; 5
2006	Paper 2	1	1; 2; 3; 4; 5
2006	Combined	1	Papers 1 & 2
2007	Paper 1	1	1; 2; 3; 4; 5
2007	Paper 2	1	1; 2; 3; 4; 5
2007	Combined	1	Papers 1 & 2

b SG

Year	Paper	Profile	Questions
2001	Paper 1	1	1; 2; 3; 4; 5
2001	Paper 2	1	1; 2; 3; 4; 5
2001	Combined	1	Papers 1 & 2
2002	Paper 1	1	1; 2; 3; 4; 5
2002	Paper 2	1	1; 2; 3; 4; 5
2002	Combined	1	Papers 1 & 2
2003	Paper 1	1	1; 2; 3; 4; 5
2003	Paper 2	1	1; 2; 3; 4; 5
2003	Combined	1	Papers 1 & 2
2004	Paper 1	1	1; 2; 3; 4; 5
2004	Paper 2	1	1; 2; 3; 4; 5
2004	Combined	1	Papers 1 & 2
2005	Paper 1	1	1; 2; 3; 4; 5
2005	Paper 2	1	1; 2; 3; 4; 5
2005	Combined	1	Papers 1 & 2
2006	Paper 1	1	1; 2; 3; 4; 5
2006	Paper 2	1	1; 2; 3; 4; 5
2006	Combined	1	Papers 1 & 2
2007	Paper 1	1	1; 2; 3; 4; 5
2007	Paper 2	1	1; 2; 3; 4; 5
2007	Combined	1	Papers 1 & 2

Note:
a Combined for comparative purposes.

Appendix 5.5 continued

I. 2001 to 2007 IEB HG (a) and SG (b)

a HG

Year	Profile	Questions
2001	1	1; 2; 3; 4; 5; 6
2001	2	1; 2; 3; 4; 5; 7
2002	1	1; 2; 3; 4; 5; 6
2002	2	1; 2; 3; 4; 5; 7
2003	1	1; 2; 3; 4; 5; 6
2003	2	1; 2; 3; 4; 5; 7
2004	1	1; 2; 3; 4; 5; 6
2004	2	1; 2; 3; 4; 5; 7
2005	1	1; 2; 3; 4; 5; 6
2005	2	1; 2; 3; 4; 5; 7
2006	1	1; 2; 3; 4; 5; 6
2006	2	1; 2; 3; 4; 5; 7
2007	1	1; 2; 3; 4; 5; 6
2007	2	1; 2; 3; 4; 5; 7

b SG

Year	Profile	Questions
2001	1	1; 2; 3; 4; 5; 6; 7
2001	2	1; 2; 3; 4; 5; 6; 8
2002	1	1; 2; 3; 4; 5; 6; 7
2002	2	1; 2; 3; 4; 5; 6; 8
2003	1	1; 2; 3; 4; 5; 6; 7
2003	2	1; 2; 3; 4; 5; 6; 8
2004	1	1; 2; 3; 4; 5; 6; 7
2004	2	1; 2; 3; 4; 5; 6; 8
2005	1	1; 2; 3; 4; 5; 6; 7
2005	2	1; 2; 3; 4; 5; 6; 8
2006	1	1; 2; 3; 4; 5; 6; 7
2006	2	1; 2; 3; 4; 5; 6; 8
2007	1	1; 2; 3; 4; 5; 6; 7
2007	2	1; 2; 3; 4; 5; 6; 8

Appendix 6.1 Summary of marks allocated to different types of questions for 1994 Biology HG question papers.

Examination Body	Total mark	Section A							Section B	Section C		
		Multiple choice	Matching columns/ diagrams/ labels	Item - statement	Termin-ology/ one word	Missing words/ labels	Incorrect / correct labels	Short questions	Short questions	Multiple choice	Data response	Essay
Transvaal	400	60^{2}						40	225			75 (75)
Cape	400	22^{2} 42^{3} 21^{3}		30^{3}					225		22 (12)	38 (48)
Natal	330	10^{2}	12^{2}			18^{1}		50	180			60 (60)
Orange Free State	300	20^{1}	7^{1}	12^{2}	13^{1}		8^{2}		180			60 (60)
HOD	300	20^{2}	10^{1}		10^{2}			40	165		55 (55)	
HOR	300	40^{2}	15^{1}		15^{1}				180			50 (50)
DET	400	160^{2}							180	$60^{2.7}$		
NEB	400	20^{2}						80	225			75 (75)
IEB	400	36^{3}			20^{2}			59	225			60 (60)

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Numbers in brackets indicate marks for which there was a choice of questions within that section.
4. All papers were three hours long.
5. See Chapter 2 for full names of examination bodies.

Appendix 6.2 Summary of marks allocated to different types of questions for 1995 Biology HG question papers.

		Section A							Section B	Section C		
Examination Body	**Total mark**	Multiple choice	Matching columns/ diagrams/ labels	Item - statement	Termin-ology/ one word	Missing words/ labels	Incorrect -correct labels	Short questions	Short questions	Multiple choice	Data response	Essay
Transvaal	400	60^{2}	10^{1}			10^{1}		20	225			75 (75)
Cape	400	45^{3}	30^{3}	15^{3} 15^{3}				10	225		16 (16)	44 (44)
Natal	330	8^{2}	12^{2}					70	180			60 (60)
Orange Free State	300	20^{1}	7^{1}	12^{2}	13^{1}		8^{2}		180			60 (60)
HOD	300	14^{2} 6^{2}	6^{1}		12^{2}			42	165		55	(55)
HOR	300	40^{2}	10^{1}		15^{1}			5	180			50 (50)
DET	400	160^{2}							180	$60^{2.7}$		(60)
NEB	400	$26^{2.9}$						74	225			75 (75)
IEB	400	36^{3}	20^{2}		12^{1}	10^{2}		37	225			60 (60)

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Numbers in brackets indicate marks for which there was a choice of questions within that section.
4. All papers were three hours long.
5. See Chapter 2 for full names of examination bodies.

Appendix 6.3 Summary of marks allocated to different types of questions for 1996 Biology HG question papers.

Examination Body	Total mark	Section A							Section B	Section C		
		Multiple choice	Matching columns/ diagrams/ labels	Item - statement	Termin-ology/ one word	Missing words/ labels	Incorrect/ correct labels	Short questions	Short questions	Multiple choice	Data response	Essay
Gauteng	300	80 [2]						20	150		50	
Western Cape	300	40 [2]	10 [1]	30 [2]	10 [1]				165		20 (25)	23 (20)
Eastern Cape	400	55 [2]	60 [2]						225		60	(60)
Northern Cape	300	20 [2]	8 [2]	12 [2]	12 [1]		8 [2]		180	60 [3]		(60)
KwaZulu-Natal	300	14 [2]	10 [2]					61	165		22	28 (50)
Free State	300	20 [2]	14 [2]	8 [2]	12 [1]		6 [2]		180	60 [3]		(60)
Mpumalanga	300	80 [2]	10 [1]		10 [1]				150		30	20 (20)
Northern Province	400	60 [2]			20 [2]			40	225			55 (55)
North West	300	60 [2]		14 [2]	14 [2]		12 [2]		150		50	(50)
IEB	400	36 [3]						89	200 (29)			75 (75)

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Numbers in brackets indicate marks for which there was a choice of questions within that section.
4. All papers were three hours long.
5. See Chapter 2 for full names of examination bodies.

Appendix 6.4 Summary of marks allocated to different types of questions for 1997 Biology HG question papers.

Examination Body	Total mark	Section A							Section B	Section C		
		Multiple choice	Matching columns/ diagrams/ labels	Item - statement	Termin- ology/ one word	Missing words/ labels	Incorrect/ correct labels	Short questions	Short questions	Multiple choice	Data re- sponse	Essay
Gauteng	300	80^{2}	10^{1}					10	150		50	
Western Cape	300	50^{2}	10^{1}	10^{2}	20^{2}				165		45	(45)
Eastern Cape	400	40^{2} 15^{3}	20^{2}		40^{2}				225			60 (60)
Northern Cape	400	45^{3}	20^{2}		30^{3}			20	225		60 (14)	(46)
KwaZulu-Natal	300	14^{2} 10^{2}	14^{2} 10^{2}	10^{2}	12^{2}			15	165		50	(50)
Free State	300	20^{2}	10^{2}	12^{2}	10^{1}		8^{2}		180	60^{3}		(60)
Mpumalanga	400	80^{2}			20^{2} 20^{2}				210		50	20 (20)
Northern Province	400	60^{2}			20^{2}			40	225			55 (55)
North West	400	60^{2}	10^{2} 10^{2}		20^{2}				225		75	(75)
IEB	400	36^{3}			20^{2}			69	200 (39)			75 (75)

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Numbers in brackets indicate marks for which there was a choice of questions within that section.
4. All papers were three hours long.
5. See Chapter 2 for full names of examination bodies.

Appendix 6.5 Summary of marks allocated to different types of questions for 1998 Biology HG question papers.

Examination Body	**Total mark**	**Section A**							**Section B**	**Section C**		
		Multiple choice	Matching columns/ diagrams	Item - statement	Termin-ology/ one word/ one word	Missing words/ labels	Incorrect/ correct labels	Short questions	Short questions	Multiple choice	Data re-sponse	Essay
Gauteng	300	80^{2}						20	150		30	20
Western Cape	300	40^{2}	10^{1}	20^{2}	20^{2}				165		45	(45)
Eastern Cape	400	40^{2} 15^{3}	20^{2}		40^{2}				225		60 (18)	(42)
Northern Cape	400	45^{3}		30^{3}	30^{3}			10	225		38 (32)	22 (28)
KwaZulu-Natal	300	12^{2}	10^{2}	10^{2} 9^{1}	12^{2}			32	165		50	(50)
Free State	300	20^{2}	12^{2}	10^{2}	10^{1}		8^{2}		180	60^{3}	(60)	
Mpumalanga	300	80^{2}	10^{1}		10^{1}				150		30	20 (20)
Northern Province	400	30^{2}	6^{2}	24^{2}	20^{2}			40	225		25	30 (55)
North West	400	60^{2}	10^{1}		10^{1}	10^{1}	10^{2}		225		75	(75)
IEB	320	24^{2}				8^{1}		68	160 (20)			60 (60) (60)

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Numbers in brackets indicate marks for which there was a choice of questions within that section.
4. All papers were three hours long.
5. See Chapter 2 for full names of examination bodies.

Appendix 6.6 Summary of marks allocated to different types of questions for 1999 Biology HG question papers.

Examination Body	**Total mark**	**Section A**							**Section B**	**Section C**		
		Multiple choice	Matching columns/ diagrams	Item - statement	Termin-ology/on e word	Missing words/ labels	Incorrect/ correct labels	Short questions	Short questions	Multiple choice	Data re-sponse	Essay
Gauteng	300	80 [2]	10 [1]					10	150		30	20
Western Cape	300	40 [2]	10 [1]	20 [2]	20 [2]				165		45	(45)
Eastern Cape	400	40 [2] 15 [3]	20 [2]		40 [2]				225		40	20 (60)
Northern Cape	400	40 [2]	20 [2]		30 [2]			25	225		42	18 (60)
KwaZulu-Natal	300	10 [2]	6 [1]		14 [2]			55	165		50	(50)
Free State	300	20 [2]	12 [2]	14 [2]	8 [1]		6 [2]		180	60 [3]		(60)
Mpumalanga	300	80 [2]	10 [1]		10 [1]				150		30	20 (20)
Northern Province	400	60 [2]			20 [2]			40	225		30	25 (55)
North West	400	60 [2]	10 [1]	5 [1]	15 [1]		10 [2]		225		45	30 (75)
IEB	320	30 [2]	20 [2]					50	160 (20)			60 (60)

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Numbers in brackets indicate marks for which there was a choice of questions within that section.
4. All papers were three hours long.
5. See Chapter 2 for full names of examination bodies.

Appendix 6.7 Summary of marks allocated to different types of questions for 2000 Biology HG question papers.

Examination Body	Total mark	Section A							Section B	Section C		
		Multiple choice	Matching columns/ diagrams	Item - statement	Termin- ology/ one word	Missing words/ labels	Incorrect/ correct labels	Short questions	Short questions	Multiple choice	Data re- sponse	Essay
Gauteng	300	80^{2}						20	150		30	20
Western Cape	300	40^{2}	10^{1}	20^{2}	20^{2}				165		45	(45)
Eastern Cape	400	40^{2}	20^{2}	15^{3}	40^{2}				225		60	(60)
Northern Cape	400	40^{2}		30^{2}	5^{1} 40^{2}				225		60	(60)
KwaZulu-Natal 1	200	10^{2}		10^{2}	10^{1}			30	105		17	18
KwaZulu-Natal 2	200	10^{2}	10^{2} 8^{2}		10^{1}			22	105		18	17
Free State	300	20^{2}	14^{2}	10^{2}	10^{1}		6^{2}		180	60^{3}		(60)
Mpumalanga	300	80^{2}	10^{1}	10^{1}					150		30	20 (20)
Northern Province	400	60^{2}	20^{2}		20^{2}			20	225			55 (55)
North West	400	60^{2}	10^{1}		10^{1} 10^{1}	10^{1}			225		75	(75)
IEB	320	18^{2}			20^{2}			62	160 (18)			60 (60)

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Numbers in brackets indicate marks for which there was a choice of questions within that section.
4. All papers were three hours long except for KwaZulu-Natal which had two question papers which were two hours each.
5. See Chapter 2 for full names of examination bodies.

Appendix 6.8 Summary of marks allocated to different types of questions for 2001 to 2007 national DoE Biology HG question papers.

Year, paper number	Total mark	Section A							Section B	Section C	
		Multiple choice	Matching columns/ diagrams	Item - statement	Termin-ology/ one word	Missing words/ labels	Incorrect/ correct labels	Short questions	Short questions	Data response	Essay
2001 Paper 1	200	14^{2}		12^{2}	12^{1}			22	105	17	18
2001 Paper 2	200	14^{2}		12^{2}	10^{1}			24	105	17	18
2002 Paper 1	200	14^{2}	10^{2}		10^{1}			26	105	17	18
2002 Paper 2	200	20^{2}	10^{2}	6^{2}	10^{1}	8^{1}	6^{2}		105	17	18
2003 Paper 1	200	12^{2}		16^{2}	10^{1}			22	105	17	18
2003 Paper 2	200	10^{2}	10^{2}	10^{2}	10^{1}			20	105	17	18
2004 Paper 1	200	14^{2}		14^{2}	5^{1}			27	105	17	18
2004 Paper 2	200	12^{2}		12^{2}	8^{1}			28	105	17	18
2005 Paper 1	200	14^{2}		10^{2}	6^{1}			30	105	17	18
2005 Paper 2	200	14^{2}		10^{2}	6^{1}			30	105	17	18
2006 Paper 1	200	14^{2}		12^{2}	7^{1}	8^{1}		19	105	17	18
2006 Paper 2	200	16^{2}		12^{2}	9^{1}			23	105	17	18
2007 Paper 1	200	14^{2}		12^{2}	6^{1}			28	105	17	18
2007 Paper 2	200	16^{2}		12^{2}		8^{1}		24	105	17	18

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. All papers were two hours long.

Appendix 6.9 Summary of marks allocated to different types of questions for 2001 to 2007 national DoE Biology SG question papers analyzed in this study.

Year, paper number	**Total mark**	**Section A**							**Section B**
		Multiple choice	Matching columns/ diagrams	Item - statement	Termin-ology/ one word	Missing words/ labels	Incorrect/ correct labels	Short questions	Short questions
2001 Paper 1	150	10^{2}	10^{2}		10^{1}			20	100
2001 Paper 2	150	10^{2}	10^{2} 10^{2}		10^{1}			10	100
2002 Paper 1	150	10^{2}	6^{1}	14^{2}	10^{1}			10	100
2002 Paper 2	150	20^{2}	10^{2}		10^{1}			10	100
2003 Paper 1	150	8^{2}	8^{2} 10^{2}		9^{1}			15	100
2003 Paper 2	150	10^{2}	10^{2}		10^{1}			20	100
2004 Paper 1	150	12^{2}	14^{2}		6^{1}		4^{2}	14	100
2004 Paper 2	150	10^{2}	12^{2}		7^{1}			21	100
2005 Paper 1	150	12^{2}	12^{2}		8^{1}			18	100
2005 Paper 2	150	16^{2}	10^{2}		6^{1}		6^{2}	12	100
2006 Paper 1	150	14^{2}	14^{2}		5^{1}			17	100
2006 Paper 2	150	14^{2}	12^{2}		8^{1}			16	100
2007 Paper 1	150	10^{2}	14^{2}		8^{1}			18	100
2007 Paper 2	150	16^{2}	12^{2}		6^{1}			16	100

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. All papers were two hours long.

Appendix 6.10 Summary of marks allocated to different types of questions for 2001 to 2007 IEB Biology HG and SG question papers.

Year, paper number, grade	**Total mark**	**Section A**							**Section B**	**Section C**	
		Multiple choice	Matching columns/ diagrams	Item ï statement	Termin-ology/ one word	Missing words/ labels	True/ false Present/ absent	Short questions	Short questions	Data response	Essay
2001 Higher Grade	300	14^{2}	20^{2}					66	140		60 (60)
2002 Higher Grade	400	18^{2}			20^{2}		20^{1}	67	200		75 (75)
2003 Higher Grade	300	20^{2}	10^{2}		10^{1}			60	140		60 (60)
2004 Higher Grade	300	20^{2}	20^{2}					60	140		60 (60)
2005 Higher Grade	300	28^{2}	20^{2}					52	140		60 (60)
2006 Higher Grade	300	20^{2}	20^{2}					60	140		60 (60)
2007 Higher Grade	300	16^{2}	20^{2}					64	140		60 (60)
2001 Standard Grade	225	20^{2}	10^{2} 16^{2}					34	125		20 (20)
2002 Standard Grade	300	20^{2}	10^{2}		10^{2}	10^{2}		50	175 (8)		25 (25)
2003 Standard Grade	225	20^{2}	10^{1}		38^{2}			12	125		20 (20)
2004 Standard Grade	225	16^{2}	10^{1}		40^{2}			14	125		20 (20)
2005 Standard Grade	225	10^{2}	10^{1} 4^{1} 10^{2}		36^{2}			10	125		20 (20)
2006 Standard Grade	225	14^{2}	10^{1}	4^{1}			10^{1}	42	125		20 (20)
2007 Standard Grade	225	12^{2}	5^{1} 7^{1} 14^{2}			8^{1} 8^{1}		26	125		20 (20)

Note: 1. The superscript next to a mark indicates the mark(s) allocated per question.
2. More than one mark in a cell means more than one sub-question of that type was found in Section A.
3. Numbers in brackets indicate marks for which there was a choice of questions within that section.
4. All papers were three hours long.
5. Both 2002 SG Section B options have identical profiles.

Appendix 6.11 Types of questions, as percentages of total marks, in the SC Biology HG question papers 1994.

Examination Body	Profile	Total marks	One sentence	Two - three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
Transvaal	1, 2	400	77.3	4.0	18.8	20.5	12.3	5.8
Cape	1	400	82.3	9.8	8.0	18.0	8.5	0.0
Cape	2	400	85.3	6.8	8.0	20.0	5.5	2.0
Natal	1	330	61.5	31.2	7.3	15.5	6.7	0.0
Natal	2	330	79.7	13.0	7.3	15.5	6.7	0.0
Orange Free State	1, 2	300	94.3	5.7	0.0	23.3	0.7	0.0
HOD	1	300	89.7	6.7	3.7	24.0	0.0	10.0
HOD	2	300	96.3	3.7	0.0	24.0	0.0	4.0
HOR	1	300	76.3	23.0	0.7	32.3	3.7	2.7
HOR	2	300	93.0	6.3	0.7	32.3	3.7	2.7
DET	1	400	91.3	8.7	0.0	27.3	4.0	1.8
NEB	1	400	62.0	15.0	23.0	42.5	12.8	5.5
NEB	2	400	62.0	15.0	23.0	23.8	12.8	5.5
IEB	1, 2	400	56.3	33.0	10.8	15.5	4.5	3.5
Mean percentage			78.5	13.2	8.3	23.1	5.8	3.1
(Std Dev)			(14.2)	(10.5)	(8.2)	(7.3)	(4.5)	(2.8)

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3)
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.12 Types of questions, as percentages of total marks, in the SC Biology HG question papers 1995.

Examination Body	Profile	Total marks	One sentence	Two - three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
Transvaal	1, 2	400	77.3	3.0	19.8	40.6	5.1	7.0
Cape	1	400	78.5	20.5	1.0	17.3	13.3	7.0
Cape	2	400	78.5	9.5	12.0	17.3	12.3	8.5
Natal	1, 2	330	74.8	3.0	22.1	10.9	15.5	5.5
Orange Free State	1, 2	300	78.7	21.3	0.0	19.7	0.7	4.0
HOD	1	300	72.7	26.0	1.3	9.7	2.7	2.3
HOD	2	300	79.0	8.3	12.7	9.7	7.7	2.3
HOR	1	300	76.7	14.3	9.0	29.0	0.0	0.0
HOR	2	300	91.7	8.3	0.0	29.0	3.3	0.0
DET	1	400	92.3	7.8	0.0	31.8	10.0	3.0
DET	2	400	80.5	19.5	0.0	21.0	6.0	3.0
NEB	1, 2	400	75.5	20.5	4.0	38.8	8.2	1.0
IEB	1, 2	400	53.3	33.0	13.8	19.8	0.0	2.3
Mean percentage			76.1	15.3	8.6	23.5	6.3	3.6
(Std Dev)			(9.8)	(10.0)	(8.5)	(10.9)	(5.3)	(2.6)

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.13 Types of questions as percentages of total marks, in the SC Biology HG question papers 1996.

Examination Body	Profile	Total marks	One sentence	Two-three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
Gauteng	1	300	82.3	14.0	3.7	40.7	4.0	2.7
Western Cape	1	300	93.3	6.7	0.0	7.7	3.7	0.7
Western Cape	2	300	91.0	9.0	0.0	8.7	5.0	2.0
Eastern Cape	1	400	83.5	15.5	1.0	24.5	1.5	2.0
Eastern Cape	2	400	96.5	2.5	1.0	24.5	1.5	4.0
Northern Cape	1	300	94.0	4.0	2.0	26.0	7.7	4.7
Northern Cape	2	300	87.0	11.0	2.0	31.0	7.7	4.7
KwaZulu-Natal	1	300	68.7	31.3	0.0	22.0	9.0	7.7
KwaZulu-Natal	2	300	85.3	14.7	0.0	22.0	9.0	7.7
Free State	1	300	97.3	2.7	0.0	33.0	4.0	0.0
Free State	2	300	81.3	15.7	3.0	39.0	10.0	2.0
Mpumalanga	1	300	91.0	7.3	1.7	28.3	4.3	0.0
Mpumalanga	2	300	84.3	14.0	1.7	28.3	4.3	0.0
Northern Province	1,2	400	85.5	13.4	1.1	28.8	4.0	7.3
Northern Province	2	400	85.5	13.4	1.1	31.3	4.0	7.3
North West	1	300	66.7	23.0	10.3	25.0	5.7	7.3
North West	2	300	59.0	30.7	10.3	13.0	1.0	7.3
IEB	1, 2, 5, 6	400	68.0	25.8	6.3	17.3	5.8	7.8
Mean percentage			81.2	15.8	3.0	23.9	5.0	4.7
(Std Dev)			(11.5)	(9.1)	(3.3)	(8.8)	(2.2)	(3.1)

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3). For example, IEB had six profiles
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.14 Types of questions, as percentages of total marks, in the SC Biology HG question papers 1997.

Examination Body	Profile	Total marks	One sentence	Two-three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
Gauteng	1	300	74.7	23.7	1.7	35.7	6.3	3.3
Western Cape	1	300	83.7	13.0	3.3	20.3	4.7	10.3
Western Cape	2	300	80.3	16.3	3.3	20.3	8.7	10.3
Eastern Cape	1, 2	400	96.5	3.5	0.0	19.5	3.8	2.0
Northern Cape	1	400	93.3	4.0	2.8	35.3	8.0	0.0
Northern Cape	2	400	94.8	2.5	2.8	35.3	4.0	0.0
KwaZulu-Natal	1	300	89.0	5.3	5.7	27.3	10.7	2.3
KwaZulu-Natal	2	300	83.3	10.3	6.3	27.3	13.3	2.3
Free State	1	300	87.3	12.7	0.0	21.7	4.7	0.0
Free State	2	300	80.3	19.7	0.0	29.7	9.7	0.0
Mpumalanga	1	400	88.3	10.3	1.5	42.8	4.5	3.0
Mpumalanga	2	400	86.3	12.3	1.5	42.8	4.5	3.0
Northern Province	1	400	86.0	13.0	1.0	22.3	10.3	5.8
Northern Province	2	400	75.5	23.5	1.0	20.8	9.0	5.8
North West	1	400	78.5	15.9	5.8	26.3	4.9	7.0
North West	2	400	62.3	34.8	3.3	19.0	4.9	2.5
IEB	1, 2	400	64.0	32.8	3.3	25.5	6.3	1.0
IEB	3, 4	400	64.0	32.8	3.3	22.0	6.3	1.0
IEB	5, 6	400	64.0	32.8	3.3	28.0	4.3	1.0
IEB	7, 8	400	64.0	32.8	3.3	24.5	4.3	1.0
Mean			77.9	19.4	2.6	26.3	6.3	2.7
(Std Dev)			(12.2)	(11.7)	(1.7)	(7.2)	(2.6)	(2.9)

Note:
1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3). For example, IEB had eight profiles.
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.15 Types of questions, as percentages of total marks, in the SC Biology HG question papers 1998.

Examination Body	Profile	Total marks	One sentence	Two-three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
Gauteng	1	300	79.0	15.3	5.7	35.0	2.0	7.7
Western Cape	1	300	83.0	16.3	0.7	20.3	0.0	10.7
Western Cape	2	300	77.0	22.3	0.7	20.3	0.0	6.3
Eastern Cape	1	400	91.0	6.5	2.5	18.3	4.3	4.3
Eastern Cape	2	400	82.0	17.0	1.0	18.3	4.3	4.3
Northern Cape	1	400	63.5	26.3	10.3	20.3	12.8	1.0
Northern Cape	2	400	74.0	20.8	5.3	20.3	12.8	1.5
KwaZulu-Natal	1	300	88.7	7.3	4.0	20.0	5.0	6.7
KwaZulu-Natal	2	300	86.0	8.7	5.3	21.3	5.0	8.7
Free State	1	300	78.7	20.3	1.0	17.7	3.3	1.0
Free State	2	300	91.0	8.0	1.0	27.0	5.3	1.0
Mpumalanga	1	300	87.0	12.3	0.7	13.0	3.7	3.3
Mpumalanga	2	300	80.3	19.0	0.7	13.0	3.7	3.3
Northern Province	1	400	76.5	20.0	3.5	23.3	3.3	2.5
Northern Province	2	400	84.0	9.5	6.5	23.3	3.3	3.3
North West	1	400	85.3	11.9	3.0	41.5	4.1	5.5
North West	2	400	69.5	27.8	3.0	27.8	4.1	5.5
IEB	1, 2, 3 4, 5, 6	320	55.0	18.4	26.6	19.7	4.4	5.9
Mean			74.2	16.5	9.3	21.7	4.5	4.9
(Std Dev)			(13.3)	(5.8)	(10.7)	6.3	2.9	2.5

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, section 5.2.1.3). For example, IEB had six profiles.
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.16 Types of questions, as percentages of total marks, in the SC Biology HG question papers 1999.

Examination Body	Profile	Total marks	One sentence	Two-three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
Gauteng	1	300	83.7	14.7	1.7	26.3	9.7	0.3
Western Cape	1	300	87.0	12.3	0.7	18.7	8.7	9.0
Western Cape	2	300	84.7	12.3	3.0	19.3	8.7	9.0
Eastern Cape	1	400	83.0	12.5	4.5	23.5	5.8	4.0
Eastern Cape	2	400	88.0	7.5	4.5	23.5	5.8	4.0
Northern Cape	1	400	86.0	11.0	3.0	29.5	7.8	5.8
Northern Cape	2	400	72.0	26.0	2.0	20.3	7.8	5.8
KwaZulu-Natal	1	300	73.3	22.0	4.7	21.0	10.0	6.7
KwaZulu-Natal	2	300	87.0	9.0	6.0	24.3	12.3	8.7
Free State	1	300	64.3	30.0	5.7	21.7	0.0	0.0
Free State	2	300	77.3	16.0	6.7	32.7	3.0	0.0
Mpumalanga	1	300	84.7	7.3	8.0	16.7	2.0	11.0
Mpumalanga	2	300	84.7	14.0	1.3	16.7	2.0	11.0
Northern Province	1	400	86.8	9.1	4.3	22.3	10.1	1.9
Northern Province	2	400	80.8	15.1	4.3	22.5	12.6	1.9
North West	1	400	88.3	10.6	1.1	30.0	3.9	8.8
North West	2	400	69.5	29.5	1.1	25.8	3.9	5.0
IEB	1, 2	320	73.1	24.7	2.2	19.1	1.3	0.0
IEB	3, 4	320	70.6	25.3	4.1	19.1	3.1	1.9
Mean			79.4	17.1	3.6	22.4	5.8	4.6
(Std Dev)			(7.5)	(7.6)	(2.0)	(4.4)	(3.8)	(3.8)

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3). For example, IEB had four profiles.
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.17 Types of questions, as percentages of total marks, in the SC Biology HG question papers 2000.

Examination Body	Profile	Total marks	One sentence	Two-three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
Gauteng	1	300	82.0	17.3	0.7	48.0	6.7	5.7
Western Cape	1	300	88.7	9.7	1.7	28.3	3.3	9.0
Western Cape	2	300	91.3	7.0	1.7	15.0	3.0	8.0
Eastern Cape	1	400	80.0	17.0	3.0	15.5	13.0	0.0
Eastern Cape	2	400	72.0	26.0	2.0	15.5	11.0	0.0
Northern Cape	1	400	99.0	1.0	0.0	27.5	0.5	5.5
Northern Cape	2	400	85.3	14.8	0.0	26.5	0.5	5.5
KwaZulu-Natal [c]	1	400	87.3	5.3	7.5	35.3	14.0	1.5
Free State	1	300	77.0	18.3	4.7	48.3	5.3	3.0
Free State	2	300	65.0	34.3	0.7	43.0	1.3	0.0
Mpumalanga	1	300	73.3	26.0	0.7	26.7	9.7	3.3
Mpumalanga	2	300	77.7	21.7	0.7	26.7	9.7	3.3
Northern Province	1	400	93.5	2.3	4.3	20.3	10.4	1.0
Northern Province	2	400	79.8	2.3	18.0	20.3	10.4	1.0
North West	1	400	74.8	25.0	0.4	30.3	6.0	0.4
North West	2	400	78.0	21.6	0.4	21.5	6.0	0.4
IEB	1, 2	320	49.1	43.4	7.5	31.3	3.8	3.8
IEB	3, 4	320	51.6	43.4	5.0	31.3	3.8	3.8
IEB	5, 6	320	47.2	45.3	7.5	31.9	3.8	3.8
IEB	7, 8	320	49.7	45.3	5.0	31.9	3.8	3.8
Mean			70.1	26.1	3.9	28.9	5.5	3.3
(Std Dev)			(17.1)	(15.9)	4.1	9.0	3.5	2.4

Note:
1. See Chapter 2 (Table 2.4) for full names of examination bodies,
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3). For example, IEB had eight profiles
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.18 Types of questions, as percentages of total marks in the national DoE and IEB SC Biology HG question papers 2001-2007.

Examination Body	Year	Profile	Total marks	One sentence	Two-three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
DoE	2001	1	400	98.5	0.3	1.3	28.8	6.3	7.5
DoE	2002	1	400	91.5	8.0	0.5	22.5	5.8	14.3
DoE	2003	1	400	90.3	8.5	1.3	35.3	8.8	11.0
DoE	2004	1	400	82.8	13.3	4.0	48.3	8.8	11.5
DoE	2005	1	400	55.5	40.3	4.3	32.3	9.8	18.0
DoE	2006	1	400	76.8	12.3	11.0	25.8	7.3	14.5
DoE	2007	1	400	66.5	27.3	6.3	19.8	0.5	15.5
Mean				80.3	15.7	4.1	30.4	6.8	13.3
(Std Dev)				(15.2)	(13.6)	(3.7)	(9.5)	3.2	3.5
IEB	2001	1, 2	300	57.0	29.3	13.7	19.3	8.0	4.7
IEB	2002	1	400	61.0	38.0	1.0	19.0	4.5	4.8
IEB	2002	2	400	61.0	19.3	19.8	19.0	4.5	4.8
IEB	2003	1	300	61.3	34.3	4.3	35.0	5.7	5.7
IEB	2003	2	300	61.3	34.3	4.3	27.0	5.7	5.7
IEB	2004	1, 2	300	56.3	37.7	6.0	21.7	13.0	6.7
IEB	2005	1, 2	300	64.0	32.3	3.7	25.3	11.0	8.3
IEB	2006	1, 2	300	59.0	32.7	8.3	20.3	16.7	3.0
IEB	2007	1, 2	300	61.3	29.3	9.3	23.0	9.0	0.0
Mean				60.0	32.0	8.0	22.8	9.7	4.7
(Std Dev)				2.6	4.8	5.0	4.4	4.1	2.6

Note:
1. Total marks of the national DoE questions paper were obtained by combining data for categories across two papers written on different days.
2. IEB question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.19 Types of questions, as percentages of total marks in the national DoE and IEB SC Biology SG question papers 2001-2007.

Examination Body	Year	Profile	Total marks	One sentence	Two-three sentences	> three sentences	Use of diagrams	Use of graphs	Use of tables
DoE	2001	1	300	96.7	0.0	3.3	37.7	7.3	3.3
DoE	2002	1	300	93.0	5.3	1.7	48.7	6.7	3.0
DoE	2003	1	300	96.0	3.7	0.3	37.0	8.3	14.3
DoE	2004	1	300	95.7	3.3	1.0	44.0	11.0	4.0
DoE	2005	1	300	75.0	16.0	9.0	42.7	6.0	5.7
DoE	2006	1	300	84.0	11.0	5.0	47.0	8.3	3.7
DoE	2007	1	300	89.3	7.0	3.7	37.3	5.0	9.3
Mean				90.0	6.6	3.4	42.0	7.5	6.2
(Std Dev)				(8.0)	(5.4)	(2.9)	(4.8)	(2.0)	(4.2)
IEB	2001	1	225	84.0	13.3	2.7	36.9	6.2	3.6
IEB	2001	2	225	92.9	4.4	2.7	36.9	6.2	3.6
IEB	2002	1, 2	300	87.7	8.3	4.0	45.3	9.0	2.7
IEB	2003	1, 2	225	81.8	16.4	1.8	35.6	3.1	0.0
IEB	2004	1	225	64.0	28.0	8.0	42.7	5.3	0.0
IEB	2004	2	225	64.0	19.1	16.9	42.7	5.3	0.0
IEB	2005	1	225	70.7	17.3	12.0	37.8	5.8	2.7
IEB	2005	2	225	70.7	26.2	3.1	37.8	5.8	2.7
IEB	2006	1, 2	225	68.0	14.7	17.3	11.1	6.7	13.8
IEB	2007	1, 2	225	53.8	27.6	18.7	38.7	1.3	4.9
Mean				73.5	17.3	9.2	35.4	5.2	3.9
(Std Dev)				(12.6)	(7.7)	(7.2)	(10.8)	(2.2)	(4.5)

Note: 1. Total marks of the national DoE questions papers were obtained by combining data for categories across two papers written on different days
2. IEB question papers have multiple profiles (see Chapter 5, Section 5.2.1.3)
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.20 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1994.

Examination Body	Profile	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Require s tables
Transvaal	1, 2	400	16.3	83.8	13.5	15.8	42.3	12.3	0.0	0.0
Cape	1	400	28.8	71.3	14.0	28.8	25.0	3.5	0.0	0.0
Cape	2	400	28.8	71.3	13.0	22.3	31.5	3.5	1.0	0.0
Natal	1, 2	330	7.0	93.0	20.3	20.9	47.3	4.5	0.0	0.0
Orange Free State	1	300	14.7	85.3	15.7	20.3	48.0	0.0	1.3	0.0
Orange Free State	2	300	14.7	85.3	17.0	27.0	40.0	0.0	1.3	0.0
HOD	1	300	8.3	91.7	23.7	40.7	22.0	0.0	5.3	0.0
HOD	2	300	8.3	91.7	22.7	30.3	35.0	0.0	3.7	0.0
HOR	1, 2	300	13.3	86.7	25.0	21.0	40.7	0.0	0.0	0.0
DET	1	400	55.0	45.0	18.3	10.3	10.3	6.0	0.0	0.0
NEB	1, 2	400	7.0	93.0	18.8	16.0	55.3	3.0	0.0	0.0
IEB	1, 2	400	9.3	90.8	14.0	21.5	49.8	1.0	3.0	1.5
Mean			19.5	80.5	18.0	21.7	40.1	3.2	1.1	0.2
(Std Dev)			(14.2)	(14.2)	(4.3)	(7.0)	(12.4)	(4.0)	(1.7)	(0.5)

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.21 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1995.

Examination Body	Profile	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
Transvaal	1, 2	400	17.5	82.5	28.5	14.8	30.0	7.9	0.0	1.5
Cape	1	400	30.8	69.3	25.8	27.0	15.0	2.0	0.0	0.0
Cape	2	400	30.3	69.8	26.8	26.0	15.0	2.0	0.0	0.0
Natal	1, 2	330	6.1	93.9	18.5	22.4	43.9	3.6	1.8	3.6
Orange Free State	1, 2	300	13.0	87.0	18.7	15.7	50.0	0.7	0.0	2.0
HOD	1	300	12.0	88.0	19.7	25.7	42.7	0.0	0.0	0.0
HOD	2	300	12.0	88.0	23.0	30.3	34.7	0.0	0.0	0.0
HOR	1	300	18.0	82.0	15.3	16.0	46.3	1.7	0.0	2.7
HOR	2	300	18.0	82.0	18.0	16.7	43.0	1.7	0.0	2.7
DET	1	300	55.0	45.0	12.7	10.8	19.0	0.0	0.0	2.8
DET	2	400	40.0	60.0	14.3	10.8	32.3	0.0	0.0	2.8
NEB	1, 2	400	6.5	93.5	11.6	29.5	46.8	2.2	0.7	2.8
IEB	1, 2	400	14.0	86.0	11.5	10.3	47.0	12.5	3.0	1.8
Mean			18.3	81.7	18.5	19.0	37.7	3.4	0.6	1.9
(Std Dev)			(13.0)	(13.0)	(6.0)	(7.0)	(11.6)	(4.1)	(1.1)	(1.2)

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.22 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1996.

ı Body	Profile	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
Gauteng	1	300	26.7	73.3	22.7	19.0	17.3	14.3	0.0	0.0
Western Cape	1	300	26.7	73.3	15.0	15.0	43.3	0.0	0.0	0.0
Western Cape	2	300	26.7	73.3	16.3	17.3	39.7	0.0	0.0	0.0
Eastern Cape	1	400	28.8	71.3	27.3	27.5	11.5	5.0	0.0	0.0
Eastern Cape	2	400	28.8	71.3	34.3	18.5	11.5	5.0	2.0	0.0
Northern Cape	1	300	13.3	86.7	26.3	21.3	39.0	0.0	0.0	0.0
Northern Cape	2	300	33.3	66.7	26.3	21.3	19.0	0.0	0.0	0.0
KwaZulu-Natal	1, 2	300	11.3	88.7	15.7	36.7	35.7	0.7	0.0	0.0
Free State	1	300	14.0	86.0	20.0	30.7	35.3	0.0	0.0	0.0
Free State	2	300	34.0	66.0	20.0	30.7	15.3	0.0	0.0	0.0
Mpumalanga	1, 2	300	31.3	68.7	21.3	24.0	14.7	3.3	2.0	3.3
Northern Province	1, 2	400	17.1	83.0	31.3	16.0	25.8	7.5	0.0	2.5
North West	1	300	24.7	75.3	24.0	19.0	24.0	6.3	2.0	0.0
North West	2	300	24.7	75.3	21.0	16.7	29.3	6.3	2.0	0.0
IEB	1, 2	400	10.5	89.5	17.8	17.0	51.5	0.8	2.5	0.0
IEB	5, 6	400	10.5	89.5	17.8	14.8	53.8	0.8	2.5	0.0
Mean			21.0	79.0	21.4	21.8	30.9	3.0	1.0	0.6
(Std Dev)			(8.9)	(8.9)	(5.8)	(7.5)	(14.5)	(3.8)	(1.1)	(1.2)

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3). For example, IEB had six profi
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.23 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1997.

Examination Body	Profile	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
Gauteng	1	300	27.0	73.0	17.7	15.0	28.7	9.7	2.0	0.0
Western Cape	1	300	25.7	74.3	15.0	32.0	25.7	0.0	1.7	0.0
Western Cape	2	300	25.7	74.3	16.3	37.0	19.3	0.0	1.7	0.0
Eastern Cape	1, 2	400	19.3	80.8	37.3	16.5	21.5	4.0	1.5	0.0
Northern Cape	1	400	19.5	80.5	41.0	17.3	16.8	5.5	0.0	0.0
Northern Cape	2	400	18.5	81.5	39.0	17.8	19.3	5.5	0.0	0.0
KwaZulu-Natal	1	300	20.7	79.3	20.7	21.0	35.7	2.0	0.0	0.0
KwaZulu-Natal	2	300	21.3	78.7	23.3	29.0	24.3	2.0	0.0	0.0
Free State	1	300	14.0	86.0	14.7	25.3	46.0	0.0	0.0	0.0
Free State	2	300	34.0	66.0	14.7	23.7	27.7	0.0	0.0	0.0
Mpumalanga	1, 2	400	22.5	77.5	33.3	15.5	22.3	5.5	0.0	1.0
Northern Province	1	400	19.3	80.8	24.0	18.5	29.8	6.8	1.8	0.0
Northern Province	2	400	19.3	80.8	23.3	19.5	31.5	5.0	1.5	0.0
North West	1	400	20.0	80.0	24.8	22.0	24.5	8.8	0.0	0.0
North West	2	400	20.0	80.0	23.3	13.3	37.5	6.0	0.0	0.0
IEB	1, 2	400	9.0	91.0	27.3	25.8	35.5	2.5	0.0	0.0
IEB	3, 4	400	9.0	91.0	21.8	28.0	35.5	5.8	0.0	0.0
IEB	5, 6	400	9.0	91.0	26.3	25.3	37.0	2.5	0.0	0.0
IEB	7, 8	400	9.0	91.0	20.8	27.5	37.0	5.8	0.0	0.0
Mean			16.8	83.2	24.8	23.1	29.7	4.1	0.5	0.1
(Std Dev)			(6.2)	(6.2)	(7.5)	(6.1)	(7.7)	(2.6)	(0.8)	(0.3)

Note:
1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3). For example, IEB had eight profiles.
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.24 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1998.

Examination Body	Profile	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
Gauteng	1	300	32.7	67.3	22.3	26.7	11.0	3.3	3.3	0.7
Western Cape	1	300	26.3	73.7	21.0	24.7	23.3	2.0	2.7	0.0
Western Cape	2	300	26.3	73.7	19.7	20.3	31.7	2.0	0.0	0.0
Eastern Cape	1	400	24.5	75.5	18.0	30.5	23.5	3.5	0.0	0.0
Eastern Cape	2	400	24.5	75.5	16.5	28.0	27.5	3.5	0.0	0.0
Northern Cape	1	400	22.8	77.3	21.3	14.8	40.3	1.0	0.0	0.0
Northern Cape	2	400	22.8	77.3	21.3	13.3	38.3	1.0	3.5	0.0
KwaZulu-Natal	1	300	13.7	86.3	16.7	30.7	36.0	3.0	0.0	0.0
KwaZulu-Natal	2	300	15.7	84.3	21.0	39.7	20.7	3.0	0.0	0.0
Free State	1	300	14.0	86.0	21.7	31.7	30.7	0.0	0.0	2.0
Free State	2	300	34.0	66.0	21.7	31.7	10.7	0.0	0.0	2.0
Mpumalanga	1, 2	300	30.0	70.0	16.0	12.3	35.7	4.0	0.0	2.0
Northern Province	1	400	16.9	83.3	31.5	14.0	25.8	10.3	2.1	0.0
Northern Province	2	400	16.9	83.3	31.5	17.0	22.8	10.3	2.1	0.0
North West	1	400	21.3	78.8	35.3	21.8	16.5	5.3	0.0	0.0
North West	2	400	19.8	80.3	29.5	12.8	35.3	2.8	0.0	0.0
IEB	1, 2, 3, 4, 5, 6	320	9.7	90.3	10.0	26.9	51.3	2.2	0.0	0.0
Mean			19.6	80.4	19.0	22.9	33.4	2.9	0.6	0.4
(Std Dev)			(8.1)	(8.1)	(6.9)	(7.4)	(13.0)	(2.7)	(1.2)	(0.8)

Note:
1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3). For example, IEB had six profiles.
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.25 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 1999.

Examination Body	Profile	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
Gauteng	1	300	33.3	66.7	14.7	16.7	24.3	9.7	1.3	0.0
Western Cape	1	300	25.0	75.0	29.3	23.7	22.0	0.0	0.0	0.0
Western Cape	2	300	25.0	75.0	33.7	28.3	13.0	0.0	0.0	0.0
Eastern Cape	1	400	20.3	79.8	28.3	23.3	24.0	4.3	0.0	0.0
Eastern Cape	2	400	20.3	79.8	27.3	22.3	26.0	4.3	0.0	0.0
Northern Cape	1	400	17.3	82.8	22.3	15.5	43.0	0.0	2.0	0.0
Northern Cape	2	400	17.3	82.8	21.3	11.8	47.8	0.0	2.0	0.0
KwaZulu-Natal	1	300	7.0	93.0	21.0	33.7	36.3	2.0	0.0	0.0
KwaZulu-Natal	2	300	7.0	93.0	23.7	42.3	25.0	2.0	0.0	0.0
Free State	1	300	18.3	81.7	12.3	24.7	41.3	0.0	0.0	3.3
Free State	2	300	38.3	61.7	12.3	24.7	21.3	0.0	0.0	3.3
Mpumalanga	1, 2	300	30.0	70.0	25.7	24.7	9.7	8.0	2.0	0.0
Northern Province	1	400	15.8	84.3	30.5	30.0	23.8	0.0	0.0	0.0
Northern Province	2	400	15.8	84.3	32.5	31.3	20.5	0.0	0.0	0.0
North West	1	400	24.0	76.0	19.3	30.8	13.5	10.3	0.0	2.3
North West	2	400	20.0	80.0	18.5	28.3	27.3	3.8	0.0	2.3
IEB	1, 2, 3, 4	320	15.6	84.4	19.1	17.5	42.8	1.3	1.3	2.5
Mean			20.4	79.7	22.6	23.3	28.6	2.7	0.7	1.0
(Std Dev)			(8.0)	(7.9)	(5.8)	(7.6)	(12.3)	(3.4)	(0.8)	(1.3)

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3). For example, IEB had four profiles.
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.26 Types of free response answers, as a percentage of total marks, in the SC Biology HG question papers 2000.

Examination Body	Profile	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
Gauteng	1	300	27.0	73.0	25.3	29.3	12.7	3.0	2.7	0.0
Western Cape	1	300	27.0	73.0	23.7	29.7	17.7	2.0	0.0	0.0
Western Cape	2	300	27.0	73.0	18.3	25.3	27.3	2.0	0.0	0.0
Eastern Cape	1	400	20.3	79.8	35.5	26.5	17.8	0.0	0.0	0.0
Eastern Cape	2	400	20.3	79.8	32.0	17.0	30.8	0.0	0.0	0.0
Northern Cape	1	400	18.0	82.0	28.3	18.0	32.8	1.5	1.5	0.0
Northern Cape	2	400	18.0	82.0	27.8	15.5	35.8	1.5	1.5	0.0
KwaZulu-Natal [c]	1	400	7.5	92.5	28.8	38.5	21.3	0.5	0.0	3.5
Free State	1	300	34.7	65.3	25.3	18.0	22.0	0.0	0.0	0.0
Free State	2	300	14.7	85.3	25.3	18.0	42.0	0.0	0.0	0.0
Mpumalanga	1, 2	300	31.7	68.3	19.3	19.7	25.7	2.0	0.0	1.7
Northern Province	1	400	23.0	77.0	26.0	25.8	14.8	8.5	0.0	2.0
Northern Province	2	400	23.0	77.0	26.0	24.3	18.3	8.5	0.0	0.0
North West	1	400	17.9	82.3	29.0	25.5	19.3	8.8	0.0	0.0
North West	2	400	17.9	82.3	26.0	18.5	33.0	5.0	0.0	0.0
IEB	1, 2 3, 4	320	5.6	94.4	24.7	17.5	47.2	5.0	0.0	0.0
IEB	5, 6 7, 8	320	5.6	94.4	25.9	19.4	44.1	5.0	0.0	0.0
Mean			17.3	82.7	25.4	20.0	32.1	3.5	0.2	0.2
(Std Dev)			(9.9)	(9.9)	(3.7)	(4.7)	(12.2)	(2.4)	(0.7)	(0.6)

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3). For example, IEB had eight profile
3. Mean and Std Dev percentage calculated using profiles per question paper.

Appendix 6.27 Types of free response answers, as percentages of total marks in the national DoE and IEB SC Biology HG question papers 2001-2007.

Examination Body	Year	Profile	Total marks	Choose correct answers	Free response answers	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
DoE	2001	1	400	13.0	87.0	26.0	37.5	22.3	0.0	1.3	0.0
DoE	2002	1	400	22.0	78.0	24.3	23.3	19.3	3.8	4.5	3.0
DoE	2003	1	400	18.3	81.8	22.0	38.0	17.0	1.8	2.0	1.0
DoE	2004	1	400	14.0	86.0	19.8	39.0	18.3	2.5	5.5	1.0
DoE	2005	1	400	16.3	83.8	13.3	44.0	19.3	3.5	2.8	1.0
DoE	2006	1	400	19.8	80.3	18.3	31.3	22.3	0.8	6.0	1.8
DoE	2007	1	400	16.8	83.3	19.5	36.0	20.3	1.5	6.0	0.0
Mean				17.1	82.9	20.5	35.0	19.8	2.0	4.0	1.1
(Std Dev)				(3.2)	(3.2)	(4.3)	(6.1)	(2.0)	(1.4)	(2.0)	(1.0)
IEB	2001	1, 2	300	13.3	86.7	22.3	17.3	39.0	7.3	0.7	0.0
IEB	2002	1, 2	400	5.0	95.0	28.0	13.3	50.8	3.0	0.0	0.0
IEB	2003	1, 2	300	6.7	93.3	22.3	26.0	40.3	2.7	0.0	2.0
IEB	2004	1, 2	300	13.3	86.7	23.0	27.7	33.3	2.7	0.0	0.0
IEB	2005	1, 2	300	17.7	82.3	18.0	22.7	37.0	3.3	1.3	0.0
IEB	2006	1, 2	300	13.3	86.7	7.3	31.3	39.7	6.3	0.0	2.0
IEB	2007	1, 2	300	12.3	87.7	12.7	32.7	35.3	7.0	0.0	0.0
Mean				11.7	88.3	19.0	23.8	39.3	4.6	0.3	0.6
(Std Dev)				(4.2)	(4.2)	(6.7)	(6.6)	(5.4)	(2.1)	(0.5)	(0.9)

Note:
1. Total marks of the national DoE questions papers were obtained by combining data for categories across two papers written on different days.
2. IEB question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Mean and Std Dev percentage calculated using profiles per question paper

Appendix 6.28 Types of free response answers, as percentages of total marks in the national DoE and IEB SC Biology SG question papers 2001-2007.

Examination Body	Year	Profile	Total marks	Choose correct answer	Free response answer	One / two terms	Short	Extended	Requires diagrams	Requires graphs	Requires tables
DoE	2001	1	300	16.7	83.3	39.7	26.0	11.3	5.0	0.0	1.3
DoE	2002	1	300	21.3	78.7	42.3	26.3	8.0	2.0	0.0	0.0
DoE	2003	1	300	22.0	78.0	40.3	28.0	5.3	4.3	0.0	0.0
DoE	2004	1	300	18.3	81.7	37.3	28.3	10.0	4.7	0.0	1.3
DoE	2005	1	300	17.0	83.0	30.3	42.3	8.0	2.3	0.0	0.0
DoE	2006	1	300	23.3	76.7	26.3	40.3	4.3	3.3	0.0	2.3
DoE	2007	1	300	22.3	77.7	28.0	33.7	6.7	6.0	0.0	3.3
Mean				19.4	80.6	34.3	31.2	7.4	4.0	0.0	1.2
(Std Dev)				(2.3)	(2.3)	(6.9)	(6.6)	(2.1)	(1.5)	(0.0)	(1.3)
IEB	2001	1, 2	225	21.3	78.7	26.2	28.4	22.2	0.9	0.9	0.0
IEB	2002	1, 2	300	14.7	85.3	22.0	26.3	37.0	0.0	0.0	0.0
IEB	2003	1, 2	225	13.8	86.2	45.8	16.9	21.8	1.8	0.0	0.0
IEB	2004	1, 2	225	14.7	85.3	46.2	18.2	19.1	0.0	1.8	0.0
IEB	2005	1, 2	225	21.8	78.2	44.4	11.6	20.9	1.3	0.0	0.0
IEB	2006	1, 2	225	22.2	77.8	18.2	42.7	16.9	0.0	0.0	0.0
IEB	2007	1, 2	225	40.0	60.0	24.4	19.6	14.7	1.3	0.0	0.0
Mean				22.7	77.3	32.2	23.0	21.5	0.8	0.4	0.0
(Std Dev)				(9.7)	(9.7)	(12.1)	(9.8)	(7.2)	(0.7)	(0.7)	(0.0)

Note: 1. Total marks of the national DoE questions papers were obtained by combining data for categories across two papers written on different days.
2. IEB question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Mean and Std Dev percentage calculated using profiles per question paper

Appendix 6.29 Percentages per topic in the 1994 SC Biology HG examinations.

Examination Body	Profile	Bio-chemistry	Photo-synthesis	Res-piration	Human nutrition	Human gaseous exchange	Population dynamics	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermo-regulation	Human circula-tion	Osmoreg-ulation	Outside syllabus
Transvaal	1	**15**	1	8	5	2	2	4	4	10	**31**	**21**			
Transvaal	2	**15**	4	8	5	**17**	2	4	4	10	**31**	3			
Cape	1	**13**	1		**19**	6	8	1	**16**	5	**14**	**18**			
Cape	2	**13**	**16**		**19**	6	8	1	**16**	5	**14**	3			
Natal	1	**11**	7	6	**20**	**12**	8		**14**	4	**18**				
Natal	2	**11**	7	6	**11**	3	8		**32**	4	**18**				
Orange Free State	1	**12**	**18**	**18**	9	5	4		9	9	**11**	6		6	
Orange Free State	2	**12**	7	8	9	9	4		9	**15**	**21**	6		2	
HOD	1	**17**	**15**	3	3	10	**12**		**11**	10	**17**	1			
HOD	2	1	**15**	3	1	10	**12**		**11**	10	**17**	**19**			
HOR	1	**11**	8	1	**12**	6	10	2	**11**	6	**21**	6			
HOR	2	**11**	8	8	**22**	1	10	2	5	6	**21**	6			
DET	1	**14**	10	**11**	6	5	10	2	**14**	7	**15**	6			
NEB	1	**17**		**21**	1	1	**11**	1	**15**	5	**19**	10			
NEB	2	**17**		**21**	1	1	**11**	1	**15**	5	**29**	1			
IEB	1	**18**	10	8	**16**	5	10	2	6	1	**20**	7			
IEB	2	**18**	10	8	1	5	10	2	6	**16**	**20**	7			

Note: *1.* See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3)
3. Bold typeface indicates percentages > 10%.
4. IEB syllabus 1996 to 2003 ï human circulation replaces population dynamics.
5. *Amoeba* and *Lumbricus* ï included in human excretion, national DoE syllabus 1994 to 2003.

Appendix 6.30 Percentages per topic in the 1995 SC Biology HG examinations.

Examination Body	Profile	Bio-chemistry	Photo-synthesis	Res-piration	Human nutrition	Human gaseous exchange	Population dynamics	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermo-regulation	Human circula-tion	Osmoreg-ulation	Outside syllabus
Transvaal	1	5	9	3	4	1	6	1	**17**	3	**31**	**20**		1	
Transvaal	2	**24**	9	3	4	1	6	1	**17**	3	**31**	2		1	
Cape	1	**18**	5	4	6	7	**15**		**12**	8	**12**	**16**			
Cape	2	**18**	5	4	**21**	7	**15**		**12**	8	**12**	1			
Natal	1	**30**	6	8	6	9	7	2	**12**	8	8	3		1	
Natal	2	**12**	6	8	6	9	7	2	**12**	8	**27**	3		1	
Orange Free State	1	5	7	8	6	**11**	6		**12**	**14**	**22**	8			
Orange Free State	2	9	7	**12**	6	**17**	6		**12**	9	**16**	5			
HOD	1	6	7	5	**13**	1		1	**18**	**22**	**22**	7		1	
HOD	2	6	7	5	**13**	1	**18**	1	**18**	3	**22**	7		1	
HOR	1	**13**	10	**13**	8	10	1	2	4	2	**21**	**17**			
HOR	2	**13**	**15**	**13**	8	10	**13**	2	4	2	**14**	7			
DET	1	**11**	**12**	10	7	5	10	2	**14**	6	**19**	3			
DET	2	**17**	6	10	7	5	10	2	**14**	6	**19**	3			
NEB	1,2	9	**19**	9	**11**	1	7	4	8	8	**13**	**11**			
IEB	1	**31**	8	3	6	6	**14**		10	8	**12**	2			
IEB	2	**16**	8	3	6	6	**14**		10	8	**28**	2			

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3)
3. Bold typeface indicates percentages > 10%.
4. IEB syllabus 1996 to 2003 ï human circulation replaces population dynamics.
5. *Amoeba* and *Lumbricus* ï included in human excretion, national DoE syllabus 1994 to 2003.

Appendix 6.31 Percentages per topic in the 1996 SC Biology HG examinations.

Examination Body	Profile	Bio-chemistry	Photo-synthesis	Res-piration	Human nutrition	Human gaseous exchange	Population dynamics	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermo-regulation	Human circula-tion	Osmoreg-ulation	Outside syllabus
Gauteng	1	**15**	8	8	9	9	6	3		**10**	**21**	8		1	
Western Cape	1	**13**	8	9	**16**	**11**	**11**	1	**11**	3	**13**	4		1	
Western Cape	2	**13**		1	**16**	**11**	**11**	1	**11**	3	**13**	**19**		1	
Eastern Cape	1	**14**	5	7	**11**	9	4	1	8	7	**27**	6		2	
Eastern Cape	2	**14**	5	7	**11**	9	**19**	1	8	7	**12**	6		2	
Northern Cape	1	10	10	**11**	**13**	2	9		**20**	7	**15**	3			
Northern Cape	2	**15**	4	**13**	**15**	2	9		**13**	9	**15**	3		3	
KwaZulu-Natal	1	**24**	**15**	6	1	5	**24**	2	**13**		9	1			
KwaZulu-Natal	2	7	**15**	6	1	5	**24**	2	**13**	**17**	9	1			
Free State	1	10	9	**12**	**12**	**12**	10		**18**	2	**13**	3			
Free State	2	8		10	**14**	**16**	10		**18**	6	**15**	4			
Mpumalanga	1	5	9	**18**	**15**	4	8	1	**12**	3	**18**	6		1	
Mpumalanga	2	5	9	**18**	8	4	8	1	**19**	3	**18**	6		1	
Northern Province	1	**15**	**15**	6	7	7	5	3	**21**	6	6	8		3	
Northern Province	2	**12**	4	6	7	7	5	3	**21**	6	**19**	8		3	
Northwest	1	**15**	**15**	**13**	6	4	6		**17**	8	10	1		5	
Northwest	2	10	**13**	3	6	4	6		**17**	8	**27**	1		5	
IEB	1	**23**	9		7	**14**			9	**12**	**11**	1	**14**		
IEB	2	**23**			7	**14**				**12**	**11**	**20**	**14**		
IEB	5	**20**	**13**		7	**14**			9	**12**	**11**	1	**14**		
IEB	6	**20**	3		7	**14**				**12**	**11**	**20**	**14**		

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3)
3. Bold typeface indicates percentages > 10%.
4. IEB syllabus 1996 to 2003 ï human circulation replaces population dynamics.
5. *Amoeba* and *Lumbricus* ï included in human excretion, national DoE syllabus 1994 to 2003.

Appendix 6.32 Percentages per topic in the 1997 SC Biology HG examinations.

Examination Body	Profile	Bio-chemistry	Photo-synthesis	Res-piration	Human nutrition	Human gaseous exchange	Population dynamics	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermo-regulation	Human circula-tion	Osmoreg-ulation	Outside syllabus
Gauteng	1	10	9	7	**12**	6	8	3	**11**	9	**17**	6		1	
Western Cape	1	**13**	5	1	**13**	4	**13**	1	**16**	**11**	**17**	5		1	
Western Cape	2	**13**	5	6	**17**	**11**	**13**	1	1	**11**	**17**	5		1	
Eastern Cape	1	**12**	6	7	**12**	**21**	9	1	9	6	**15**	3		1	
Eastern Cape	2	**22**	6	7	**12**	6	9	1	9	**11**	**15**	3		1	
Northern Cape	1	**15**		9	9	10	10	4	9	6	**13**	**16**		1	
Northern Cape	2	**15**	**12**	9	9	10	**13**	4	9	6	**13**	1		1	
KwaZulu-Natal	1	5	10	7	8	6	**13**	5	**15**	**17**	**13**	1			
KwaZulu-Natal	2	9	10	7	**21**	6	**13**	5	**15**		**13**	1			
Free State	1	**11**	8	**19**	**16**	1	4		**16**	8	**13**	5		1	
Free State	2	4	**11**	**16**	9	2	8		**19**	8	**16**	8		1	
Mpumalanga	1	7	**12**	8	**20**	6	4		10	10	**20**	4		1	
Mpumalanga	2	7	**12**	8	**15**	6	4		10	**12**	**20**	7		1	
Northern Province	1	**17**	9	4	8	2	6	2	6	8	**23**	**13**		2	
Northern Province	2	**17**	3		8	2	6	2	2	8	**37**	**13**		2	
Northwest	1	10	10	9	5	7	7	5	**30**	8	10	1			
Northwest	2	7	3	9	5	7	**26**	5	**21**	8	10	1			
IEB	1, 3 5, 7	1	5		7	**19**			**14**	**14**	**20**	6	**16**		
IEB	2, 4 6, 8	1	5		**26**				**14**	**14**	**20**	6	**16**		

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Bold typeface indicates percentages > 10%.
4. IEB syllabus 1996 to 2003 ï human circulation replaces population dynamics.
5. *Amoeba* and *Lumbricus* ï included in human excretion, national DoE syllabus 1994 to 2003.

Appendix 6.33 Percentages per topic in the 1998 SC Biology HG examinations.

Examination Body	Profile	Bio-chemistry	Photo-synthesis	Res-piration	Human nutrition	Human gaseous exchange	Population dynamics	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermo-regulation	Human circula-tion	Osmoreg-ulation	Outside syllabus
Gauteng	1	**13**	7	8	**15**	6	6	3	9	7	**21**	4			
Western Cape	1	**14**	10	3	**12**		10	1	**12**	**12**	**22**	3		1	
Western Cape	2	**14**	6	3	**12**			1	**12**	**12**	**37**	2		1	
Eastern Cape	1	**32**	8	7	**13**	4	8	3	5	4	**11**	7			
Eastern Cape	2	**17**	8	7	**13**	4	8	3	5	**19**	**11**	7			
Northern Cape	1	**24**	2	6	**12**	**12**	10	2	2	**11**	**15**	6		1	
Northern Cape	2	**18**	2		**12**	8	10	2	**17**	**11**	**15**	6		1	
KwaZulu-Natal	1	6	**12**		10	10	**11**	5	**11**	10	**17**	9			
KwaZulu-Natal	2	8	**12**	**14**	10	10	**11**	5	**11**	10		9			
Free State	1	**15**	6	10	5	10	10		**18**	6	8	7		4	
Free State	2	**15**	8	7	7	10	**12**		**22**	7	3	4		4	
Mpumalanga	1	**16**	1	1	**16**	5	9	7	9	6	**13**	**14**		3	
Mpumalanga	2	**16**	1	1	**16**	5	9	7	**15**	1	**13**	**12**		3	
Northern Province	1	10	4	8	**12**	10	8	1	9	9	**22**	9		1	
Northern Province	2	**21**	4	8	**12**	6	8	1	**12**	9	**18**	3		1	
Northwest	1	**11**	**11**	**15**	**11**	7	7		**14**	9	**13**	3		1	
Northwest	2	7	9	9	**21**	**16**	7		7	9	**13**	3		1	
IEB	1, 4	**11**	3	3	**13**	**17**			**19**	**19**		6	9		
IEB	2, 5	**11**	3	3	**13**	**17**			**19**		**19**	6	9		
IEB	3, 6	**11**	3	3	**13**	**17**			**19**			6	9		**19**

Note: *1.* See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Bold typeface indicates percentages > 10%.
4. IEB syllabus 1996 to 2003 ï human circulation replaces population dynamics.
5. *Amoeba* and *Lumbricus* ï included in human excretion, national DoE syllabus 1994 to 2003.

Appendix 6.34 Percentages per topic in the 1999 SC Biology HG examinations.

Examination Body	Profile	Bio-chemistry	Photo-synthesis	Res-piration	Human nutrition	Human gaseous exchange	Population dynamics	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermo-regulation	Human circula-tion	Osmoreg-ulation	Outside syllabus
Gauteng	1	6	10	5	**12**	10	8		**15**	7	**21**	3		4	
Western Cape	1	**14**	3	9	**17**	9	9	1	**17**	6	**14**	1		1	
Western Cape	2	**19**	7	2	**18**	3	9	2	**18**	6	**14**	1		1	
Eastern Cape	1	**11**	8	8	**12**	**13**	8	1	**16**	5	**19**	1			
Eastern Cape	2	**11**	8	5	**14**	6	8	1	**16**	5	**21**	6			
Northern Cape	1	**16**	8	2	4	7	**11**	2	9	**15**	**15**	9		2	
Northern Cape	2	**16**	8	2	**19**	7	**11**	2	9	2	**15**	9			
KwaZulu-Natal	1	7	10	10	0	4	**14**	3	**17**	**15**	**17**	3			
KwaZulu-Natal	2	**12**	10	10	10	4	**14**	3		**16**	**17**	3			
Free State	1	8	**13**	9	10	1	7		**28**	6	**13**	2		3	
Free State	2	6	3	5	**13**	3	7		**35**	7	**13**	5		3	
Mpumalanga	1	**16**	8	4	7	6	**14**	2	**12**	6	**20**	6			
Mpumalanga	2	**16**	**13**	4	9	6	**14**	2	**12**	6	**13**	6			
Northern Province	1	**15**	8	5	5	4	**15**	3	5	2	**29**	6		3	
Northern Province	2	**22**	**13**	5	8	4	**15**	3	5	2	**15**	6		3	
Northwest	1	8	**14**	**12**	9	7	8		**16**	9	**12**	6			
Northwest	2	10	**20**	10	9	7	8		9	11	**12**	6			
IEB	1, 3	**13**		8	9	**13**			**19**	10	**16**	1	**13**		
IEB	2, 4	**13**		8	9	**13**				10	**16**	**19**	**13**		

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Bold typeface indicates percentages > 10%.
4. IEB syllabus 1996 to 2003 ï human circulation replaces population dynamics.
5. *Amoeba* and *Lumbricus* ï included in human excretion, national DoE syllabus 1994 to 2003.

Appendix 6.35 Percentages per topic in the 2000 SC Biology HG examinations.

Examination Body	Profile	Bio-chemistry	Photo-synthesis	Res-piration	Human nutrition	Human gaseous exchange	Population dynamics	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermo-regulation	Human circula-tion	Osmoreg-ulation	Outside syllabus
Gauteng	1	10	7	10	10	5	8	3	**12**	8	**17**	6		4	
Western Cape	1	**12**	2	6	**15**	9	**12**		**17**	8	**19**			1	
Western Cape	2	**12**	2	6	**15**	9	**12**		2	8	**19**	**15**		1	
Eastern Cape	1	**19**	5	6	**15**	7	9	4	8	7	**15**	6		1	
Eastern Cape	2	10	5	6	10	**22**	9	4	8	7	**15**	6		1	
Northern Cape	1	**14**	9	6	**13**	9	**14**		9		**12**	**14**		1	
Northern Cape	2	**14**	9	6	**13**	9	**11**		9	**15**	**12**	2		1	
KwaZulu-Natal	1	8	5	4	**11**	**12**	10	3	**13**	**12**	**16**	6			
Free State	1	**11**	**17**	8	**12**	5	7	2	**12**	**11**	4	6		4	
Free State	2	9	**15**	7	7	5	7		10	8	**23**	4		4	
Mpumalanga	1	**13**	10	4	6	5	**15**	2	**14**	8	**14**	9			
Mpumalanga	2	**13**	10	4	**12**	5	**15**	2	**14**	8	**14**	3			
Northern Province	1	**20**	9	6	4	4	6	1	**13**	6	**20**	8		5	
Northern Province	2	**20**	9	6	**18**	4	6	1	**13**	6	**12**	3		5	
Northwest	1	**12**	**15**	**13**	10	5	6	1	**11**	10	**15**	4		1	
Northwest	2	**14**	10	10	10	**12**	6	1	8	10	**15**	4		1	
IEB	1, 3	10	**13**	9	10	**13**			9	**11**	**14**	3	6	2	
IEB	2, 4	10	4		**19**	**22**			9	**11**	**14**	3	6	2	
IEB	5, 7	10	**13**	**13**	10	9			9	**11**	**14**	3	6	2	
IEB	6, 8	10	4	3	**19**	**19**			9	**11**	**14**	3	6	2	

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Bold typeface indicates percentages > 10%.
4. IEB syllabus 1996 to 2003 ï human circulation replaces population dynamics.
5. *Amoeba* and *Lumbricus* ï included in human excretion, national DoE syllabus 1994 to 2003
6. KwaZulu-Natal, Paper 1 and paper 2 combined.

Appendix 6.36 Percentages per topic in the2001-2007 national DoE and IEB SC Biology HG examinations.

Examination Body	Year	Profile	Bio-chemistry	Photo-synthesis	Res-piration	Human nutrition	Human gaseous exchange	Population dynamics	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermo-regulation	Human circula-tion	Osmo-regulation	Outside syllabus
DoE	2001	1	**11**	6	6	10	8	9	1	**15**	8	**20**	6			
IEB	2001	1	8		6	**28**	8			**12**	**12**	**13**	9	6		
IEB	2001	2	8		6	**18**	3			**22**	**22**	**13**	4	6		
DoE	2002	1	8	10	4	10	10	10	2	**16**	**11**	**15**	6			
IEB	2002	1	**13**	3	7	**16**	6			**24**	9	7	8	10		
IEB	2002	2	**13**	3	7	**16**	6			5	9	**25**	8	10		
DoE	2003	1	**12**	7	6	9	8	9	2	**15**	**11**	**17**	5			
IEB	2003	1	8	0	8	**14**	8			9	**12**	**20**	**11**	10		
IEB	2003	2	8	**20**	8	**14**	8			9	**12**	10	1	10		
DoE	2004	1	**11**	5	7	4	**15**	9	2	**14**	10	**20**	4			
IEB	2004	1	**11**	**16**	1	8	**14**	**11**		10	**12**	9	8			
IEB	2004	2	**11**	6	1	8	4	**11**		10	**12**	**29**	8			
DoE	2005	1	9	7	7	**11**	8	9	2	**15**	**11**	**17**	5			
IEB	2005	1	**14**	2	2	**11**	8	9		9	**24**	**13**	8			
IEB	2005	2	**14**	**22**	2	**11**	8	9		9	4	**13**	8			
DoE	2006	1	7	7	3	**12**	**12**	9	3	**15**	**11**	**17**	6			
IEB	2006	1	5	1	6	**32**	6	7		8	8	**22**	5			
IEB	2006	2	5	1	6	**12**	6	**27**		8	8	**22**	5			
DoE	2007	1	5	7	6	**15**	8	9	2	**16**	**11**	**17**	5			
IEB	2007	1	10	6	7	5	6	**13**		**17**	10	**22**	4			
IEB	2007	2	10	6	7	5	**16**	**13**		7	10	**22**	4			

Note: *1.* See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Bold typeface indicates percentages > 10%.
4. IEB syllabus 1996 to 2003 ï human circulation replaces population dynamics.
5. *Amoeba* and *Lumbricus* ï included in human excretion, national DoE syllabus 1994 to 2003
6. National DoE Paper 1 and paper 2 combined.

Appendix 6.37 Percentages per topic in the 2001-2007 national DoE and IEB SC Biology SG examinations.

Examination Body	Year	Profile	Bio-chemistry	Photo-synthesis	Res-piration	Human nutrition	Human gaseous exchange	Population dynamics	Plant hormones	Plant water relations	Human excretion	Human co-ordination	Thermo-regulation	Human circula-tion	Osmo-regulation	Outside syllabus
DoE	2001	1	**12**	8	5	**11**	7	7	2	**12**	**11**	**20**	5			
IEB	2001	1	7	4	2	**14**	**13**			**22**	6	**13**	2	**16**		
IEB	2001	2	7	**13**	2	**14**	**13**			**22**	6	**13**	2	7		
DoE	2002	1	**14**	7	6	6	8	8	1	**15**	**12**	**18**	5			
IEB	2002	1	**16**	8	1	**12**	**15**			7	8	**17**	4	**12**		
IEB	2002	2	8	8	1	**12**	**15**			7	8	**17**	**12**	**12**		
DoE	2003	1	9	7	5	**14**	5	9	2	**15**	**11**	**14**	9			
IEB	2003	1	**20**	4	7		9			9	**22**	**15**	8	6		
IEB	2003	2	**20**	4	7					9	**22**	**24**	8	6		
DoE	2004	1	**11**	5	8	9	8	9	2	**15**	**11**	**16**	6			
IEB	2004	1	**11**			**13**	**18**	**11**		6	**11**	**22**	7			
IEB	2004	2	**11**	4	4	4	**18**	**11**		6	**11**	**22**	7			
DoE	2005	1	8	7	5	**12**	9	9	2	**16**	**11**	**16**	6			
IEB	2005	1	**11**	**12**		**11**		**11**		**11**	**11**	**25**	9			
IEB	2005	2	**11**	**12**		**11**	9	**11**		**11**	**11**	**25**	0			
DoE	2006	1	**14**	6	7	7	7	9	2	**14**	**11**	**18**	4			
IEB	2006	1	5	8		**26**		7		**14**	**11**	**20**	8			
IEB	2006	2	5	8		**17**		7		**14**	**20**	**20**	8			
DoE	2007	1	10	9	4	10	8	9	2	**15**	**11**	**17**	5			
IEB	2007	1	**12**	7	8	**12**	10	9		8	**11**	**12**	**11**			
IEB	2007	2	3	7	8	**12**	10	9		**17**	**11**	**12**	**11**			

Note: *1.* See Chapter 2 (Table 2.4) for full names of examination bodies.
2. Some question papers have multiple profiles (see Chapter 5, Section 5.2.1.3).
3. Bold typeface indicates percentages > 10%.
4. IEB syllabus 1996 to 2003 ï human circulation replaces population dynamics.
5. *Amoeba* and *Lumbricus* ï included in human excretion, national DoE syllabus 1994 to 2003
6. National DoE Paper 1 and paper 2 combined.

Appendix 6.38 Percentage weightings of PET categories of cognitive demand in the 1994 SC Biology HG question papers.

Examination Body	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
Transvaal	1, 2	400	52	9	25	12	3
Cape	1	400	53	6	24	9	9
Cape	2	400	66	6	13	9	7
Natal	1	330	30	8	41	0	21
Natal	2	330	48	8	22	0	21
Orange Free State	1, 2	300	59	0	37	1	2
HOD	1	300	52	3	34	1	10
HOD	2	300	46	2	42	0	10
HOR	1, 2	300	73	1	24	1	2
DET	1	400	89	3	8	0	1
NEB	1, 2	400	47	5	23	0	26
IEB	1, 2	400	32	4	40	4	21

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.39 Percentage weightings of PET categories of cognitive demand in the 1995 SC Biology HG question papers.

Examination Body	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
Transvaal	1, 2	400	78	9	3	2	9
Cape	1	400	62	10	12	5	11
Cape	2	400	60	10	13	5	12
Natal	1, 2	330	53	10	16	2	18
Orange Free State	1, 2	300	62	1	27	5	5
HOD	1	300	71	0	11	2	16
HOD	2	300	54	2	12	2	29
HOR	1	300	76	3	16	4	2
HOR	2	300	69	4	21	6	0
DET	1, 2	400	95	0	4	0	1
NEB	1	400	33	6	38	0	22
NEB	2	400	33	6	21	17	22
IEB	1, 2	400	45	6	24	1	25

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.40 Percentage weightings of PET categories of cognitive demand in the 1996 SC Biology HG question papers.

Examination Body	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
Gauteng	1	300	66	16	6	0	13
Western Cape	1	300	75	3	12	8	3
Western Cape	2	300	75	3	16	2	4
Eastern Cape	1	300	91	4	1	1	3
Eastern Cape	2	300	85	7	2	1	6
Northern Cape	1	300	76	3	14	4	3
Northern Cape	2	300	75	3	14	5	3
KwaZulu-Natal	1	300	56	4	24	2	14
KwaZulu-Natal	2	300	45	4	35	2	14
Free State	1	300	82	0	12	2	4
Free State	2	300	67	1	16	6	10
Mpumalanga	1, 2	300	85	1	4	7	3
Northern Province	1, 2	400	72	8	8	3	9
Northwest	1	300	70	6	4	3	18
Northwest	2	300	70	6	9	3	13
IEB	1, 2, 5, 6	400	49	9	22	3	18

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.41 Percentage weightings of PET categories of cognitive demand in the 1997 SC Biology HG question papers.

Examination Body	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
Gauteng	1	300	50	12	19	7	11
Western Cape	1	300	83	4	6	2	5
Western Cape	2	300	85	4	4	2	5
Eastern Cape	1, 2	400	85	5	8	2	0
Northern Cape	1	400	78	5	2	3	13
Northern Cape	2	400	80	6	3	3	9
KwaZulu-Natal	1	300	65	10	7	11	7
KwaZulu-Natal	2	300	57	12	8	11	12
Free State	1	300	86	0	14	0	0
Free State	2	300	82	1	14	0	3
Mpumalanga	1, 2	400	74	7	9	4	6
Northern Province	1	400	63	10	13	3	11
Northern Province	2	400	69	8	10	3	11
Northwest	1	400	57	9	13	4	18
Northwest	2	400	70	6	6	4	14
IEB	1, 2, 3, 4	400	61	7	11	10	11
IEB	5, 6, 7, 8	400	58	10	11	10	11

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.42 Percentage weightings of PET categories of cognitive demand in the 1998 SC Biology HG question papers.

Examination Body	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
Gauteng	1	300	64	10	12	1	14
Western Cape	1	300	71	4	11	5	10
Western Cape	2	300	77	1	9	3	10
Eastern Cape	1	400	60	2	28	2	8
Eastern Cape	2	400	65	2	25	2	6
Northern Cape	1	400	58	5	28	0	9
Northern Cape	2	400	60	4	31	0	6
KwaZulu-Natal	1	300	70	8	12	3	7
KwaZulu-Natal	2	300	64	8	12	4	12
Free State	1	300	85	1	7	5	2
Free State	2	300	79	1	7	6	7
Mpumalanga	1, 2	300	80	4	7	5	4
Northern Province	1	400	70	9	8	0	13
Northern Province	2	400	67	9	8	3	13
Northwest	1	400	67	5	18	2	7
Northwest	2	400	75	3	14	2	6
IEB	1, 2	320	51	5	23	12	9
IEB	3	320	44	5	23	3	26
IEB	4,5	320	54	5	19	12	9
IEB	6	320	47	5	19	3	26

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.43 Percentage PET weightings of categories of cognitive demand in the 1999 SC Biology HG question papers.

Examination Body	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
Gauteng	1	300	61	15	16	3	6
Western Cape	1	300	63	8	15	7	6
Western Cape	2	300	64	8	14	7	6
Eastern Cape	1	400	46	5	28	9	12
Eastern Cape	2	400	57	5	18	9	12
Northern Cape	1	400	46	6	39	7	3
Northern Cape	2	400	59	5	27	7	3
KwaZulu-Natal	1	300	62	7	21	1	10
KwaZulu-Natal	2	300	52	9	27	1	11
Free State	1	300	68	2	19	7	5
Free State	2	300	58	2	26	10	5
Mpumalanga	1, 2	300	68	7	17	5	3
Northern Province	1	400	64	3	24	5	4
Northern Province	2	400	63	5	24	5	4
Northwest	1	400	50	11	25	3	10
Northwest	2	400	64	5	18	3	10
IEB	1, 2	320	48	4	32	7	9
IEB	3, 4	320	48	4	29	7	12

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.44 Percentage weightings of PET categories of cognitive demand in the 2000 SC Biology HG question papers.

Examination Body	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
Gauteng	1	300	69	8	11	9	3
Western Cape	1	300	59	4	18	7	12
Western Cape	2	300	63	3	17	7	10
Eastern Cape	1	400	69	3	14	8	7
Eastern Cape	2	400	63	3	14	15	6
Northern Cape	1	400	58	6	32	1	5
Northern Cape	2	400	69	4	23	0	4
KwaZulu-Natal	1	400	61	2	20	7	11
Free State	1	300	69	1	21	2	7
Free State	2	300	81	1	15	1	2
Mpumalanga	1, 2	300	66	1	15	8	10
Northern Province	1, 2	400	68	9	15	2	7
Northwest	1	400	62	9	17	7	5
Northwest	2	400	70	6	13	6	5
IEB	1, 2, 3, 4	320	46	6	30	8	10
IEB	5, 6, 7, 8	320	49	6	27	8	10

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. KwaZulu-Natal Paper 1 and Paper 2 combined.

Appendix 6.45 Percentage weightings of categories of PET cognitive demand in the national DoE and IEB SC Biology HG question papers, 2001-2007.

Year	Examination Body	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
2001	National	1	400	52	10	24	1	14
2001	IEB	1, 2	300	45	7	21	15	12
2002	DoE	1	400	46	17	20	0	18
2002	IEB	1, 2	400	53	8	25	8	6
2003	DoE	1	400	48	11	25	0	17
2003	IEB	1, 2	300	42	7	28	13	10
2004	DoE	1	400	41	11	23	1	24
2004	IEB	1, 2	300	43	7	26	4	20
2005	DoE	1	400	41	11	13	14	21
2005	IEB	1, 2	300	42	7	26	0	25
2006	DoE	1	400	39	13	26	6	16
2006	IEB	1, 2	300	40	10	26	3	21
2007	DoE	1	400	48	9	21	0	23
2007	IEB	1, 2	300	42	12	22	0	23

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. National DoE Paper 1 and Paper 2 combined.

Appendix 6.46 Percentage weightings of PET categories of cognitive demand in the national DoE and IEB SC Biology SG question papers, 2001-2007.

Year	Examination Body	Profile	Total marks	Memorize	Perform procedure	Explain	Analyze	Apply
2001	DoE	1	300	67	10	20	0	4
2001	IEB	1, 2	225	64	5	27	3	2
2002	DoE	1	300	72	4	19	0	5
2002	IEB	1, 2	300	63	5	16	10	7
2003	DoE	1	300	64	10	18	0	8
2003	IEB	1, 2	225	77	4	16	0	3
2004	DoE	1	300	60	7	26	1	7
2004	IEB	1, 2	225	75	4	19	0	2
2005	DoE	1	300	65	6	14	1	14
2005	IEB	1, 2	225	82	4	10	0	4
2006	DoE	1	300	65	6	23	0	6
2006	IEB	1, 2	225	63	4	15	0	18
2007	DoE	1	300	62	10	18	0	9
2007	IEB	1, 2	225	73	3	12	1	11

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. National DoE Paper 1 and Paper 2 combined.

Appendix 6.47 Relationship between BOK and DOK in the 1994 SC Biology HG question papers.

Examination Body	Profile	Total marks	BOK	DOK
Transvaal	1	400	5.56	0.17
Transvaal	2	400	6.10	0.17
Cape	1	400	7.08	0.22
Cape	2	400	7.43	0.20
Natal	1	330	7.78	0.26
Natal	2	330	5.81	0.26
Orange Free State	1	300	8.55	0.03
Orange Free State	2	300	8.44	0.03
HOD	1	300	7.74	0.13
HOD	2	300	7.38	0.11
HOR	1	300	9.02	0.03
HOR	2	300	7.49	0.03
DET	1	400	9.24	0.00
NEB	1	400	6.53	0.34
NEB	2	400	5.30	0.34
IEB	1	400	7.78	0.33
IEB	2	400	7.87	0.33
Policy			None	**0.67**

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.48 Relationship between BOK and DOK in the 1995 SC Biology HG question papers.

Examination Body	**Profile**	**Total marks**	**BOK**	**DOK**
Transvaal	1	400	5.56	0.11
Transvaal	2	400	5.21	0.11
Cape	1	400	8.42	0.19
Cape	2	400	7.44	0.22
Natal	1	330	7.14	0.26
Natal	2	330	7.78	0.26
Orange Free State	1	300	8.30	0.11
Orange Free State	2	300	8.82	0.11
HOD	1	300	6.49	0.21
HOD	2	300	7.05	0.46
HOR	1	300	7.57	0.06
HOR	2	300	8.91	0.06
DET	1	400	8.71	0.00
DET	2	400	8.25	0.00
NEB	1	400	8.99	0.28
NEB	2	400	8.99	0.65
IEB	1	400	6.31	0.35
IEB	2	400	6.74	0.35
Policy			None	**0.67**

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.49 Relationship between BOK and DOK in the 1996 SC Biology HG question papers.

Examination Body	Profile	Total marks	BOK	DOK
Gauteng	1	300	8.49	0.15
Western Cape	1	300	9.14	0.12
Western Cape	2	300	7.65	0.06
Eastern Cape	1	300	7.46	0.04
Eastern Cape	2	300	9.16	0.07
Northern Cape	1	300	8.04	0.08
Northern Cape	2	300	8.56	0.09
KwaZulu-Natal	1	300	6.00	0.18
KwaZulu-Natal	2	300	6.98	0.19
Free State	1	300	8.57	0.06
Free State	2	300	7.95	0.19
Mpumalanga	1	300	8.16	0.12
Mpumalanga	2	300	7.82	0.12
Northern Province	1	400	8.74	0.13
Northern Province	2	400	8.26	0.13
Northwest	1	300	8.67	0.26
Northwest	2	300	6.97	0.18
IEB	1	400	7.31	0.25
IEB	2	400	6.33	0.25
IEB	5	400	7.69	0.25
IEB	6	400	6.88	0.25
Policy			None	**0.67**

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.50 Relationship between BOK and DOK in the 1997 SC Biology HG question papers.

Examination Body	Profile	Total marks	BOK	DOK
Gauteng	1	300	9.82	0.22
Western Cape	1	300	8.25	0.08
Western Cape	2	300	8.32	0.08
Eastern Cape	1	400	8.43	0.02
Eastern Cape	2	400	8.29	0.02
Northern Cape	1	400	9.29	0.18
Northern Cape	2	400	9.60	0.13
KwaZulu-Natal	1	300	8.87	0.21
KwaZulu-Natal	2	300	7.81	0.29
Free State	1	300	7.77	0.00
Free State	2	300	8.15	0.03
Mpumalanga	1	400	7.76	0.11
Mpumalanga	2	400	8.44	0.11
Northern Province	1	400	7.96	0.16
Northern Province	2	400	5.12	0.15
Northwest	1	400	7.04	0.28
Northwest	2	400	6.80	0.22
IEB	1, 3	400	6.93	0.27
IEB	2, 4	400	5.86	0.27
IEB	5, 7	400	6.93	0.25
IEB	6, 8	400	5.86	0.25
Policy			None	**0.67**

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.51 Relationship between BOK and DOK in the 1998 SC Biology HG question papers.

Examination Body	Profile	Total marks	BOK	DOK
Gauteng	1	300	8.64	0.17
Western Cape	1	300	7.61	0.17
Western Cape	2	300	4.89	0.15
Eastern Cape	1	400	6.54	0.10
Eastern Cape	2	400	8.78	0.09
Northern Cape	1	400	7.47	0.10
Northern Cape	2	400	7.84	0.06
KwaZulu-Natal	1	300	9.38	0.11
KwaZulu- Natal	2	300	9.73	0.19
Free State	1	300	9.47	0.08
Free State	2	300	8.46	0.15
Mpumalanga	1	300	8.96	0.10
Mpumalanga	2	300	8.35	0.10
Northern Province	1	400	8.62	0.15
Northern Province	2	400	7.94	0.19
Northwest	1	400	9.15	0.10
Northwest	2	400	8.28	0.09
IEB	1, 2, 4, 5	320	7.05	0.26
IEB	3, 6	320	7.05	0.40
Policy			None	**0.67**

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.52 Relationship between BOK and DOK in the 1999 SC Biology HG question papers.

Examination Body	Profile	Total marks	BOK	DOK
Gauteng	1	300	8.72	0.09
Western Cape	1	300	8.29	0.15
Western Cape	2	300	7.44	0.15
Eastern Cape	1	400	8.29	0.26
Eastern Cape	2	400	8.11	0.26
Northern Cape	1	400	8.94	0.10
Northern Cape	2	400	8.11	0.10
KwaZulu-Natal	1	300	7.98	0.12
KwaZulu-Natal	2	300	8.19	0.14
Free State	1	300	6.99	0.13
Free State	2	300	5.75	0.17
Mpumalanga	1	300	8.50	0.09
Mpumalanga	2	300	9.19	0.09
Northern Province	1	400	6.79	0.10
Northern Province	2	400	8.07	0.10
Northwest	1	400	9.30	0.16
Northwest	2	400	9.04	0.16
IEB	1	320	7.68	0.20
IEB	2	320	7.54	0.20
IEB	3	320	7.68	0.24
IEB	4	320	7.54	0.24
Policy			None	**0.67**

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.53 Relationship between BOK and DOK in the 2000 SC Biology HG question papers.

Examination Body	Profile	Total marks	BOK	DOK
Gauteng	1	300	10.21	0.14
Western Cape	1	300	7.73	0.23
Western Cape	2	300	8.10	0.20
Eastern Cape	1	400	8.90	0.17
Eastern Cape	2	400	8.73	0.26
Northern Cape	1	400	8.79	0.06
Northern Cape	2	400	9.14	0.04
KwaZulu-Natal [c]	1	400	9.42	0.21
Free State	1	300	9.81	0.11
Free State	2	300	8.56	0.04
Mpumalanga	1	300	9.28	0.21
Mpumalanga	2	300	8.95	0.21
Northern Province	1	400	8.13	0.09
Northern Province	2	400	8.32	0.09
Northwest	1	400	9.13	0.14
Northwest	2	400	9.46	0.12
IEB	1, 3, 5, 7	320	9.56	0.22
IEB	2, 4	320	7.15	0.22
IEB	6, 8	320	7.80	0.22
Policy			None	**0.67**

Note: 1. See Chapter 2 (Table 2.4) for full names of examination bodies.
2. KwaZulu-Natal combination of Paper 1 and Paper 2

Appendix 6.54 Relationship between BOK and DOK in the 2001-2007 national DoE and IEB SC Biology HG question papers

Year	Examination Body	Profile	Total marks	BOK	DOK
2001	DoE	1	400	8.77	0.18
2002	DoE	1	400	9.44	0.22
2003	DoE	1	400	9.26	0.20
2004	DoE	1	400	8.40	0.33
2005	DoE	1	400	9.36	0.54
2006	DoE	1	400	9.15	0.27
2007	DoE	1	400	8.96	0.29
Policy	**National**		**400**	**9.38**	**0.67**
2001	IEB	1	300	6.94	0.38
2001	IEB	2	300	6.50	0.38
2002	IEB	1	400	7.66	0.17
2002	IEB	2	400	7.34	0.17
2003	IEB	1	300	8.26	0.29
2003	IEB	2	300	8.40	0.29
2004	IEB	1	300	8.90	0.32
2004	IEB	2	300	6.71	0.32
2005	IEB	1	300	7.52	0.33
2005	IEB	2	300	7.91	0.33
2006	IEB	1	300	5.68	0.32
2006	IEB	2	300	6.30	0.32
2007	IEB	1	300	7.85	0.30
2007	IEB	2	300	7.93	0.30
Policy	**IEB**			None	**0.67**

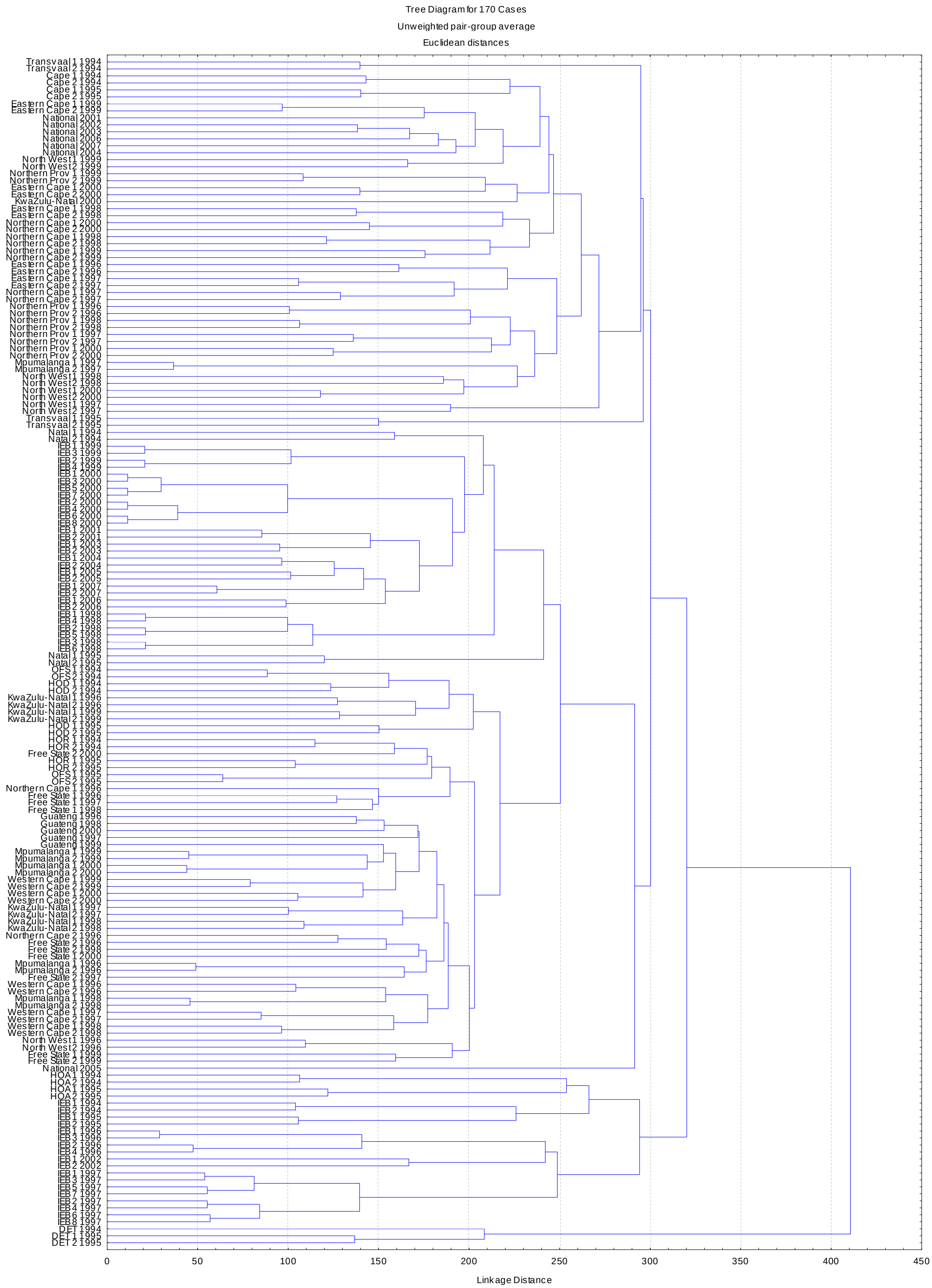
Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.55 Relationship between BOK and DOK in the 2001-2007 national DoE and IEB SC Biology SG question papers.

Year	Examination Body	Profile	Total marks	BOK	DOK
2001	DoE	1	300	8.97	0.04
2002	DoE	1	300	8.85	0.05
2003	DoE	1	300	9.56	0.09
2004	DoE	1	300	9.54	0.08
2005	DoE	1	300	9.42	0.18
2006	DoE	1	300	9.07	0.07
2007	National	1	300	9.37	0.11
Policy	**DoE**		**300**	**9.38**	**0.20 to 0.25**
2001	IEB	1	225	7.33	0.05
2001	IEB	2	225	7.59	0.05
2002	IEB	1	300	8.18	0.20
2002	IEB	2	300	8.66	0.20
2003	IEB	1	225	7.16	0.03
2003	IEB	2	225	6.00	0.03
2004	IEB	1	225	7.10	0.02
2004	IEB	2	225	7.75	0.02
2005	IEB	1	225	7.19	0.04
2005	IEB	2	225	7.19	0.04
2006	IEB	1	225	6.28	0.22
2006	IEB	2	225	6.76	0.22
2007	IEB	1	225	10.06	0.14
2007	IEB	2	225	9.11	0.14
Policy	**IEB**			None	**0.25**

Note: See Chapter 2 (Table 2.4) for full names of examination bodies.

Appendix 6.56 Cluster analysis of the profiles of SC Biology HG question papers analyzed in this study.



Appendix 6.57 Content categories 2005 and 2006 HG and SG Paper 1 – Summary of the performance level at which content categories become absent from students scripts. For example, a B means that this content was not passed at symbol C and below, i.e., only A and B candidates obtained a pass in the category.

A. 2005 HG

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	FF	*	F	GG	F
Photosynthesis	GG	F	F	G	D
Respiration	E	E FF	D	D	B
Human nutrition	FF	GG	B	D	E
Human gaseous exchange	F		D	H	
Population dynamics	G	H		*	C

2006 HG

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	F				D
Photosynthesis	E	*	E	GG	C
Respiration	FF			GG	
Human nutrition	F	H			F
Human gaseous exchange	F	E	E		H
Population dynamics	FF	E	E		FF

B. 2005 SG

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	GG	H			F
Photosynthesis	FF				
Respiration	F		E		
Human nutrition	FF		D		
Human gaseous exchange	FF		E	D	
Population dynamics	H	*	A		G

2006 SG

	Memorize	Procedure	Explain	Analyze	Apply
Biochemistry	F	GG	F		
Photosynthesis	E		F		
Respiration	F		*		
Human nutrition	E				
Human gaseous exchange	G		C		
Population dynamics	H	*	*		F

Note: 1. * means that this category was passed by students of all symbols of student performance.
2. Pass in HG = E and above; pass in SG = F and above.
3. Blank cells indicate no questions.

Appendix 6.58 Content categories 2005 and 2006 HG and SG Paper 2 – Summary of the performance level at which content categories become absent from the analysis of students scripts. For example, a B means that this content was not passed at symbol C and below, i.e., only A and B candidates obtained a pass in the category.

A. **2005 HG**

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	F			E	C
Plant water relations	G	H	F	FF	E
Human excretion	G	H	D	GG	E
Human co-ordination	F		E	C	E
Thermo-regulation	FF		D	D	

2006 HG

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	G		F		
Plant water relations	F	*	E		E
Human excretion	GG	GG	E		D
Human co-ordination	G		E	F	D
Thermo-regulation			E	FF	E

B. **2005 SG**

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	H		*		
Plant water relations	G	G	E		F
Human excretion	FF		C		C
Human co-ordination	FF		GG		G
Thermo-regulation	F	GG	H		

2006 SG

	Memorize	Procedure	Explain	Analyze	Apply
Plant hormones	FF				
Plant water relations	G	E	E		E
Human excretion	GG		FF		FF GG
Human co-ordination	GG	H	D		H
Thermo-regulation	E		F		F

Note:
1. * means that this category was passed by students of all symbols of student performance.
2. Pass in HG = E and above; pass in SG = F and above.
3. Blank cells indicate no questions.

Appendix 6.59 Questions with mean percentage achieved greater than 80% or less than 20% in the 2005 national DoE SC HG and SG Biology examinations and PET category.

Year	Paper	Category	Question	Max. mark	Explanation	PET
2005	HG1	>80%	2.1.1	2	Identify liver and pancreas in diagram	A
			1.4.2	1	Identify main part of pie chart using key	B
			1.5.3	1	Name reagent testing for starch	A
			4.2.2	3	Use given mark-recapture equation	B
			4.1.3	1	Choose which species, A or B, more tolerant to humidity from table and diagram	D
		<20%	3.2.7	4	Explain different experimental results to those given (uses table)	E
			4.3.3	3	Explain different experimental results to those given (diagram and table)	E
			4.2.3	4	Explain experimental results (table)	E
			4.3.2	4	Explain experimental results(diagram and table)	E
			2.1.9	4	Explain why active absorption necessary for products of carbohydrate and protein digestion	C
2005	HG2	>80%	2.2.2	2	Read value from graph	B
			2.3.1	1	Choose solution, A or B, with highest water potential	C
		<20%	4.1.3	3	Explain significance of blood flow in seal flipper	C
			1.6.2	2	Explain answer to Q1.6.1 (22.7%) which was experimental results	E
			4.2.4 iii	5	Links two different sections of work, hormones and thermoregulation via a diagram.	D
			1.5.1	2	Identify iris and suspensory ligaments in the eye BUT is a front view of the eye and almost all texts show the lens and suspensory ligaments using a lateral view of the eye	A
			2.1.1	1	Describe experimental results using a potometer but the diagram of potometer given is not the kind used in most textbooks	A
			4.2.4 v	2	Links two different sections of work, hormones and thermoregulation via a diagram. Depends on answer to Q 4.2.4 iv (27.4%) which depends on answer to Q 4.2.4 I (29.6%) where students could have given correct answer with no understanding.	D
			5.1.3 ii	2	Need to consult two diagrams, same section of work, to get to answer. Answer depends on Q 5.1.3 (37.8%)	D
			2.1.3	4	Question asks to explain how reliability could be improved when using the potometer. Problematic because what was being asked was ‘how to ensure that results using photometer are reliable’ if the memorandum is consulted.	E
			5.1.2 i	3	Providing an explanation for unlearned observation	E
			5.1.6 ii	2	Requires an explanation for answer to Q 5.1.6 (i) (47.5%) where the student could select one of four answers from a table without understanding.	E
2005	SG1	>80%	1.1.3	2	Name reagent testing for starch	A
			2.2.1	3	Identify liver, stomach and small intestine from diagram	A
			1.3.4	2	Match term with definition	A

Appendix 6.59 continued

Year	Paper	Category	Question	Max. mark	Explanation	PET
2005	SG1	>80%	5.3.1	3	Use given mark-recapture equation	B
			5.2.4	2	Read a value from a graph	B
		<20%	1.5.1	1	Identify small intestine from a *schematic* diagram rather than the more realistic diagrams that appear in most of the textbooks	A
			2.2.6	6	Depends on an identification made in Q 2.2.1 (89.6%) but requires student to link structure and function	C
			1.2.7	1	Give the name of a described structure (biochemistry)	A
			4.1.3	1	Depends on the answer to Q 4.1.2 (32.5%), problem may have been in the use of ñfood/substrateò wrt to cellular respiration	A
			1.5.4	2	Depends on the answer to Q 1.5.3 (48.9%) and then recalling the function of amino acids once they have been absorbed	A
			5.3.2	4	Explain learned experimental results	C
			1.2.8	1	Give the name of a described process (biochemistry)	A
2005	SG2	>80%	1.1.2	2	Choose correct answer	A
			1.1.1	2	Choose correct answer	A
		<20%	3.2.3	3	Providing and explanation for unlearned observation	E
			2.1.1	3	Calculate a rate from a graph	B
			2.2.5	3	Explain unlearned experimental results	E
			5.2.4	5	Explain counter current blood flow wrt to thermoregulation in dolphins. Candidates probably tried to link this to the graph used in the previous question (10.7%)	A
			4.3.2	2	State a learned difference but candidates might not have understood what ñtarget organò meant	A
			3.1.5.2	4	Depends on Q 3.1.5.1 (58.5%) but candidates could not describe the adaptations of the structure	A
			2.1.5	2	Candidates cannot give two learned advantages	A
			3.1.4	2	Explain the importance of microvilli in the nephron	C
			3.2.2	3	Explain unlearned observation	E
			2.2.3	2	Question asks to explain how reliability could be increased when performing an osmosis experiment. Problematic because what was being asked was ôhow to ensure that results of an osmosis experiment are reliableò if the memorandum is consulted.	E
			5.2.3	1	Depends on the answer to Q 5.2.2 (54.6%) . The line in the graph was not joined to 0 and this may have been problematic	B
			3.1.3	4	Depends on answer to Q 3.2.2 (24.0%) and requires the student to manage two processes which occur in opposite directions wrt a cell	C
			2.1.3	1	Explanation of experimental procedure	C

Appendix 6.60 Questions with mean percentage achieved greater than 80% or less than 20% in the 2006 national DoE SC HG and SG Biology examinations and PET category.

Year	Paper	Category	Question	Max. mark	Explanation	PET
2006	HG1	>80%	1.5.2	2	Reading value from a graph	B
			3.2.6	2	Reading a value from a graph	A
			1.5.1	1	Recalling term	A
			5.1.2	2	Providing an explanation for unlearned phenomenon	E
			1.3.1	2	Link term and statement	A
			1.1.1	2	Recognize a function	A
		<20%	3.1.1	2	Required a pre-grade 12 understanding of ñsolventò and or to understand the described process of chromatography	E
			2.1.3	3	Some candidates could name the substance but could not explain their reasoning	E
2006	HG2	>80%	5.1.7	2	Explain experimental procedure in unlearned experiment but answer was generic	E
			5.1.4	1	Choose which species, A or B, more suited to windy conditions from graph	B
		<20%	4.2.1	6	Explain structural adaptations	C
			4.1.5	3	Explain possible results in unlearned experiment	E
			3.4	3	Explain reason for differences in structure	C
			5.1.6	4	Explain experimental results	E
			2.4.3	2	Explanation for a structural design	C
2006	SG1	>80%	1.4.2	3	Name the lungs, diaphragm and trachea on a diagram	A
			5.3.2	3	Extract three words from passage	A
			5.2.1	3	Use given mark-recapture equation	B
			1.3.3	2	Match item with statement	A
		<20%	4.3.1	2	Compare two *schematic* diagrams	A
			4.3.2	2	Depends on the answer to Q 4.3.1 (13.5%)	
			1.5.4	1	Required student to remember that tissues are made up of cells	A
2006	SG2	>80	3.2.3a	1	Choose between conditions	E
			1.3.6	2	Match term with statement	A
			4.1.2	2	Give the function of two parts shown in a diagram	A
			3.2.1	1	Identify the pinna in a diagram	A
		<20	1.2.8	1	Required to remember that blood is a tissue	A
			1.4.4	2	Give results of the control in a learned experiment	A
			3.3.2	3	Unlearned animal. Depends on the answer to Q 3.3.1 (59.8%) where candidates choose between A and B as an answer.	E
			2.1.4	4	Describe structural adaptations	C
			5.2.3	8	Depends on answer to Q 5.2.2 (30.0%) to explain reasons for the answer.	C
			1.4.6	2	Question asks to explain how reliability could be improved when investigating phototropism. Problematic because what was being asked was ôhow to ensure that results are reliableô if the memorandum is consulted.	C
			2.1.3	3	Depends on answer to Q 2.1.2 (39.1%) to explain the function of a part.	C
			1.3.2	2	Link an item and a statement	A

Appendix 7.1 Series of possible research questions generated by this study.

1. How do we decide what should be examined in the NSC Life Sciences examination? Why is the content of only one year, Grade 12, examined in the NSC Life Sciences examination?

2. What is the optimal format of a NSC Biology examination question paper and the optimal length/time of a NSC Biology question paper?

3. What was the value of a mark in NSC Life Sciences examinations? Can a value be assigned to one mark? Should the value of one mark be tied to the length of an examination? How do the total marks of similar examinations affect their assumed equivalence?

4. How should marks be allocated in the NSC Life Sciences examinations? Could the marks allocated to a particular question be used to differentiate between candidates with different levels of competency in the NSC Life Sciences examinations?

5. How equivalent should SC Biology question papers have been with respect to content? How equivalent should NSC Life Sciences examinations be with respect to content?

6. How do we best define and recognize HOCS in a task? What proportion of an NSC question paper should comprise HOCS questions? How can the coherence of an answer in a NSC Life Sciences examination be rewarded? Do students solve a task in the same way that examiners think that they solve the specified task?

7. What makes a Biology question more complex than another Biology question? What makes a Biology question more difficult than another Biology question?

8. How do the structural characteristics of a question paper affect the influence of content standards of a question paper on student performance? Should content standards (topic and cognitive demand) have a third dimension (delivery [for teaching], structural aspects of the question paper, for example, presentation of questions and answers [for assessment])?

9. How does the language of a question, the complexity of a question and the use of non-text in a question influence how Biology students answer the question? Is the relationship between these factors the same, or different, for students who are required to read questions and write answers in their home language versus and those students who read questions and write answers in a non-home language?

10. Could the cut-scores in each of the 2005 and 2006 HG and SG SC Biology examinations have been set differently, so as to yield consistent PLDs between the two years? If the cut-scores between symbols could be set differently for these examinations, how might the cut-score-methodology apply to the NSC Life Sciences examinations? Would the performance standards then tell us more about a student's ability than the aggregate mark?

11. Could different combinations of the structural aspects and the content standards (topics and performance expectations) explain better relationships, if any, of Paper 1 and Paper 2, in the two years applicable? If so, how might the NSC Paper 1 and Paper 2 be set to maximize the reliability of the results obtained by students writing both question papers? How might the 'value' of a mark be consistently used when setting NSC Life Sciences question papers?

12. How similar, or different, are the questions in NSC Life Sciences examinations (which have replaced the SC Biology examinations and have no differentiation into HG and SG) to the HG and SG questions of the pre-2008 years?