

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

**JUDGING ESSAYS: FACTORS THAT INFLUENCE  
MARKERS**

**By**

**Lorna Olckers OLCLOR001**

**A minor dissertation submitted in partial fulfillment of the requirements for the award  
of  
Degree of Master of Philosophy (in Higher Education Studies)**

**Department of Education**

**University of Cape Town**

**2006**

**Declaration**

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

**Signature** Signature removed

**Date** 13. 2. 06

## TABLE OF CONTENTS

Acknowledgements	Page 2
Abstract	Page 3
Chapter 1 Introduction	Page 5
Chapter 2 Problem	Page 8
Chapter 3 Theory	Page 19
Chapter 4 Methodology	Page 32
Chapter 5 Data Analysis	Page 39
Chapter 6 Discussion	Page 58
Chapter 7 Conclusion	Page 67
Appendix 1 BP/BHP course outcomes	Page 69
Appendix 2 Example of BP/BHP essay question	Page 70
Appendix 3 Example of marking criteria	Page 72
Appendix 4 Letter of consent	Page 73
Appendix 5 Questionnaire	Page 74
Appendix 6 Instructions for markers	Page 75
Appendix 7 Comparative study: overall results	Page 77
Appendix 8 Comparative study: individual results	Page 78
References	Page 92

## ACKNOWLEDGEMENTS

I owe thanks to so many colleagues, friends and family members for helping to make this thesis possible.

### **I would like to say special thanks to:**

- Suellen Shay - my supervisor, mentor, guide and pillar of strength for guiding, encouraging and supporting me through what she warned me would be a “slippery process”
- The BP/BHP facilitators and students of 2004 and 2005 who so willingly participated in the study
- Trevor Gibbs for believing in me and encouraging me to rise to the challenge
- Rodney Ehrlich, my Head of Department, for giving me the time to complete this study
- Gill Faris for so smoothly taking on the role of course convenor so that I could take advantage of two months of study leave
- Rauf Sayed for sharing his expertise in statistics
- The Emerging Researchers Group for funding the process
- Carolyn Butler for her superb editing skills and supportive comments
- The Owen family for lending me their glorious space in Pringle Bay as an escape and opportunity for totally immersing myself in the study data
- My mother Bente for willingly stepping in and providing extra meals, school lifts, encouragement and support
- My colleague and friend Elmi for putting up with my whining, and now telling me I should do a PHD
- Finally, and most importantly, my husband Stephen and children Matthew and Sarah for encouraging and supporting me and giving me the space to complete this long process

## ABSTRACT

### JUDGING ESSAYS: FACTORS THAT INFLUENCE MARKERS.

Question: What underlying factors influence markers in their decision-making about essay assignment marks, and what are the implications for assessment validity?

This thesis is an exploration of assessment with particular focus on the marking of essay assignments and the validity or soundness of that process. It began with concern by markers of two first year Health Sciences courses that the validity or soundness of their interpretations was being compromised because they were marking their own students' essays – students with whom they were working closely in small groups. The decision was taken to no longer allow markers to grade their own students. Whether this was the correct decision extended into a more complex exploration of what influences markers in their decision-making processes, and the implications for validity.

A comparative study of marks assigned by markers who knew the students as opposed to the marks assigned by a second marker who did not know the students revealed no significant differences. What became clear, however, from the qualitative data taken from the interviews, reflective writing and observation of moderation sessions was that assessment of essay assignments is a socially situated interpretative process. Issues of difference and particularly context influence markers as they go about making their judgements.

An interpretive theoretical perspective, drawing strongly on the work of Hans-Georg Gadamer, is used to explain how textual and contextual factors influence markers. The relationship between the whole and the parts is illustrated as concentric circles of influence, including an inner-most circle of textual factors which is about the written product itself and the writer. Contextual factors are represented by three additional concentric circles of potential influence including the individual marker, the marking team or community of practice and the outer circle of the educational ethos. Validity of the assessment process is dependent on acknowledging and working with all these contexts because ultimately the meaning and understanding of the whole is about the parts. Evidence from the data suggests that there no easy technical solution to strengthening validity. Validity is a process and

central to the process is a commitment to opportunities for dialogue and debate around emerging differences of perspective.

University of Cape Town

# CHAPTER 1: INTRODUCTION

## 1.1 Focus

Assessment is a key element of the formal educational process that ultimately drives learning. Assessment is concerned with “the quality of teaching as well as the quality of learning: it involves us in learning from our student’s experiences, and is about changing ourselves as well as our students” (Ramsden, 1992, p 182). Inextricably linked to assessment is validity. Validity is concerned with the soundness of interpretations (Nitko, 2001) made by assessors as they go about the task of grading their students’ work. This study is an exploration of the underlying factors that influence assessors and the soundness of their interpretations.

## 1.2 Background

As a lecturer in the School of Public Health and Family Medicine at the University of Cape Town it is my responsibility to convene two first year courses for Health Science students. The focus of these courses is on professionalism with the aim of developing students into “Integrated Health Professionals”. A more detailed description of this context will be presented in Chapter 2.

My journey in academic life began some eight years ago when I was first given the responsibility of lecturing a portion of a course and assessing students’ work. All too often I had the sensation of being out of my depth. In one of my early experiences of marking I asked a colleague for assistance with grading essays. He told me that the allocation of marks is really a “gut response”, and that it comes with experience. He referred to essays that he considered to be deserving of a first class pass as having the “wow” factor! This and other experiences like it only served to confirm my previous personal suspicions that assessment is a highly subjective activity. As a student, I had frequently found myself trying to decipher exactly what success or failure was about. The mark on my paper all too often left me with questions about what it really meant, and how the assessor had come to that particular mark. What was it that influenced the markers?

Through 2000 and 2001, I became increasingly involved in the design of curricula for Health Science students at the University of Cape Town. When I was appointed to the position of course convener of the new first year courses “Becoming a Professional” and “Becoming a Health Professional”, I decided to enroll in a series of courses offered by the Centre for Higher Education Development (CHED) as I had no formal background in education. Thus my learning journey continued, combining education and curricula design with work responsibilities. The challenges related to the marking of essay assignments together with my own prior experiences motivated my interest in further pursuing a study related to assessment.

### **1.3 Problem**

This study began with my need to answer questions related to the courses I convene, but progressed to the more challenging question of what it is that influences markers in their decision-making. This question brought to mind other questions such as:

- Is it the contents of the essay alone or do other factors come into play when assessors allocate marks?
- If there are other influences, what are they and what do they mean in terms of the final mark?
- What is the impact of all these factors for the validity of the assessment?

### **1.4 Overview**

- Chapter 2 begins with the central problem of this study. This includes a detailed description of the background and context of the study as well as a review of relevant international literature that has explored similar challenges.
- Chapter 3, the Theory Chapter, draws on the work of Hans-Georg Gadamer, Etienne Wenger and others to outline the theoretical frame of this study. This is illustrated through a diagram of concentric circles of both textual and contextual influences on assessors.
- The range of qualitative and quantitative methodologies employed in this study is described in Chapter 4.
- Chapter 5 presents the data that emerged from the study in terms of the concentric circles of potential influence as described in the theory chapter.
- Chapter 6, the Discussion Chapter, analyses the findings of this study in detail.

- Chapter 7, the Conclusion, explores the possible implications for practice that have arisen from this study.

University of Cape Town

## CHAPTER 2: PROBLEM

### 2.1 Introduction

This chapter elaborates on how and why this study came to focus on the factors that influence markers in their decision-making processes and the implications of these influences for the validity of the assessment process. It outlines the context of the problem area by describing the background to this research and other studies that have informed my own conclusions.

### 2.2 Background

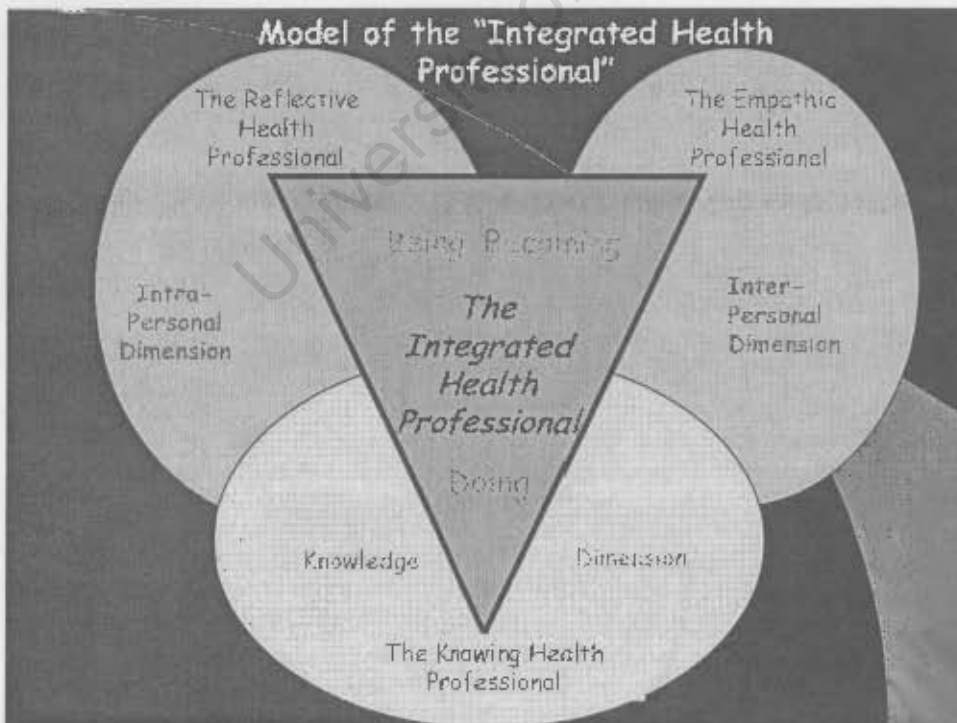
The Health Sciences Faculty at the University of Cape Town has undergone significant change in the past few years. Most significant has been the adoption of a “new” or “revised” MBChB curriculum in 2002. This shift has resulted in a rigorous process of evaluation, review and redesign of educational outcomes and activities in line with educational theory. The assessment process has been recognised as a key area of focus, and assessment activities have been evaluated and changed to ensure that there is alignment between outcomes, activities and assessment tasks.

Validity, a term once foreign to the staff, is now recognised as central to the implementation of a good, defensible MBChB curriculum. Validity, according to McMillan, “is a characteristic that refers to the appropriateness of the inferences, uses and consequences that result from the test or other method of gathering information.... Validity is concerned with the inferences, not the test itself...” and “a measure is not simply valid or invalid, it is valid to some degree” for example high, low or moderate (1997, p 55). Like McMillan, Nitko focuses on the appropriateness of the inferences made by assessors by defining validity as “the soundness of your interpretations and uses of students’ assessment results” (2001, p 36). Reliability, which Nitko sees as one dimension or one criterion for validity, is specifically concerned with consistency in marking. Reliability, or consistency, is most often equated with the idea of fairness as described by both students and assessors.

## 2.3 “Becoming a Professional” (BP) and “Becoming a Health Professional” (BHP)

### 2.3.1 Overview

The courses “Becoming a Professional” (BP) and “Becoming a Health Professional” (BHP), of which I am the convener, form part of the core faculty curriculum for all first year Health Science students at the University of Cape Town. The emphasis is on professionalism with the aim of developing students into “Integrated Health Professionals” by focusing not only on the generation of knowledge, but also on the development of interpersonal and intrapersonal or self-reflective skills, referred to as the knowing, empathic and reflective domains. The diagram below illustrates how each of these domains is seen as equally important in the being, becoming and doing of professionalism within BP/BHP. This model relates to outcomes-based education, particularly to the framework outlined by Harden et al (1999) in terms of moving from competency to meta-competency. The student develops from being able to do something through gaining technical knowledge (the knowing health professional), is then able to do it right within the context of self and others (the empathic health professional) and finally improves through further self awareness and development (the reflective health professional).



“Becoming a Professional” (BP) and “Becoming a Health Professional” (BHP) are integrated in terms of themes, principles and outcomes, but are separated by the shift from the more generic (BP) in the first semester to the context of health (BHP) in the second semester. Outcomes for the two courses (see Appendix 1) are framed within the two broad themes of interpersonal skills development and the Primary Health Care approach, with the former being the focus of BP and the latter the focus of BHP.

### **2.3.2 Students**

The student group that participates in BP/BHP is made up largely of the previous year’s matriculants. Students come from all over South Africa with a small number coming from countries such as Botswana and Mauritius. They are in the Health Sciences Faculty to study Medicine, Physiotherapy, Speech Therapy, Audiology and Occupational Therapy. BP and BHP are the only courses to date that bring all these disciplines together as a multi-professional grouping. The total number of students per year is approximately three hundred and fifty.

### **2.3.3 Course Methodology**

On joining BP and BHP, the students are divided into groups of approximately twelve students with each group being lead by a facilitator. Learning in BP/BHP takes place largely through these facilitator-lead small groups where activities are of a mostly experiential nature. The experiential activities are in turn supported by prescribed readings, journal writing, group discussion, and visits to community sites and health services as well as formal lectures. Deep as opposed to surface learning is required (Cole, 1990 in Biggs, 1996), which encourages students to “integrate knowledge, understanding and communication skills and ... to support their arguments with evidence” (Lockett and Sutherland, 2000, p 115).

### **2.3.4 Group facilitators**

The BP/BHP small group facilitators are graduates from a variety of disciplines, ranging from social work and psychology to teaching, nursing, physiotherapy and even law. They are selected on the basis of their enthusiasm for working with first year students, experience in working with small groups, openness to a more experiential methodology and a willingness to learn. Participation in intensive training is compulsory and on-going, and through this activity facilitators form a tight and supportive bond. They become a team or community of

practice (Wenger, 1998). The team encourages an atmosphere of empathy and this is reflected in the pedagogical ethos of the courses.

#### **2.4 Assessment in BP/BHP**

Assessment in BP/BHP is both formative and summative. Forty percent of the marks for BP/BHP are purely summative in nature, summing up the student's learning over the year (Light and Cox, 2001). These summative assessments take the form of written end of course examinations.

Sixty percent of the marks for each course are derived from what is referred to as in-course performance based assessments. These are both summative and formative in nature because they count for marks, but provide opportunities for extensive feedback to assist students in the learning process (Light and Cox, 2001). These assessments take the form of essay assignments, interview practicals and group presentations.

Essay assignments account for most of the in-course performance based assessments. The value of essay assignments, according to Nitko (2001), is that they assess higher order thinking as well as writing skills, not simply rote learning. Logical argument can therefore be assessed and this skill is seen as central to the students' development as Integrated Health Professionals in BP/BHP. For the purposes of this thesis, I have chosen to focus on the essay assessments.

The BP/BHP essay assignments are criterion-referenced, rather than norm-referenced, as is required of outcome-based education (Luckett and Sutherland, 2000), in that the marking criteria are made available to the markers as well as the students. (See Appendix 2 for an example of a BP essay question and Appendix 3 for an example of marking criteria). Performance is judged against the criteria, which are clear and make the process of assessment transparent (Shepard, 2000). Evidence for meeting the criteria is explicit in the instructions, and students are given grades for these assessments as well as extensive written and oral feedback.

## 2.5 Problem

In 2002 and 2003 all BP and BHP in-course performance based essay assessments were graded by the students' own group facilitators, marked by a second marker and then moderated in a session in which any difficulties were discussed, and consensus finally reached. This use of internal moderation is recognised as a useful strategy for reducing inconsistencies (Luckett and Sutherland, 2000). Despite the rigorous nature of the 2002/2003 assessment process, a number of difficulties were clearly evident. One difficulty was that marks assigned by the students' own facilitator and those of the second markers were often quite different. Second markers challenged the grades assigned by the students' own facilitators. They felt that these markers were "subjective", that is, it appeared that markers who were assessing their own students were bringing information to the assessment process that was not seen to be relevant to the assessment task. One example was the students' own facilitators wanting to be able to reward "progress" in their students which was not evident in the assessment itself, but may have been evident in the facilitator-student small group interaction. Also, facilitators who knew that their students were having a challenging time, be it academically, socially or emotionally, found it difficult to separate that knowledge from the assessment performance when assigning a mark. They were reluctant to add to their student's already difficult situation by assigning them poor grades. From the second markers' perspective these interpretations or inferences were seen as inappropriate, and a threat to the validity of the assessment. Students' own facilitators concurred that they found it hard to be "objective" and make sound judgements about their own students' work because they were effectively bringing information to the assessment process that was not seen to be relevant to the assessment task itself. This was seen as a threat to the validity of the assessment process.

Linked to the previous difficulty was the problem that some facilitators found the actual essay marking process very difficult. They struggled to shift roles from group facilitator to assessor of students' performance. This added significant stress to the job for some of the small group facilitators.

Some of the markers also struggled to give detailed written feedback to the students either because they found it difficult to quantify their concerns or because they found the task time-

consuming. The written feedback was therefore often not qualitatively consistent across markers.

A significant amount of time was put into the training of facilitators as to how to assess written work. This was found to be extremely resource intensive in terms of it being time-consuming and expensive. With the natural turnover in staff in part-time positions, training had to be repeated and the new markers typically took longer than the more experienced facilitators did to mark the essays.

In terms of the validity of scores, students themselves questioned the assigned marks in terms of their fairness after comparing marks with their friends. As course convenor, I sometimes found it difficult to defend the facilitators' marks, as they were clearly not always consistent across markers. The 2002 and 2003 BP/BHP marking system and the resulting difficulties brought into question issues around the validity of the BP/BHP assessment process.

In 2004 I introduced a new system of marking in response to the difficulties described above. I appointed an assessment team that was a subgroup of the BP/BHP facilitator team, with a particular interest in and commitment to assessing written work. In order to better manage the assessment process in the context of my own increasing workload, I appointed one of the assessment team members as the assessment co-ordinator, and two other members as her assistants. The assessment co-ordinator and her two assistants had the responsibility of designing the essay questions and marking criteria, and overseeing the assessment and moderation processes. This was all done in close consultation with me, as course convenor, as I remained the person ultimately responsible for the implementation, assessment and evaluation of BP/BHP.

The assessment team, under the guidance of the assessment co-ordinator and her two assistants, was made responsible for marking all written assessments. The markers were not allowed to mark their own students' work. Concerns about perceived "subjectivity" and possible bias were seen to be addressed by the fact that the markers were not allowed to assess their own students' essay assignments. The logic behind this decision was that if the markers did not know the students they were marking, they would not be able to bring

personal knowledge to the marking experience. Students were told about this decision in order to make the process as transparent, fair and free of bias as possible.

Those facilitators who had found the marking process too stressful and challenging were able to choose not to be part of the marking team. The expense and time-consuming nature of the previous system was also addressed through the setting up of a more consistent team of markers who saw themselves as committed to the process for at least one full year. Input on how to assess did not have to be repeated with each assessment, and this proved to be more time and cost-effective.

In response to the concern about adequate feedback, a “sandwich technique” of pointing out positive aspects of the student’s work, enumerating needed improvements and ending with a global positive comment was introduced. Ramsden (1992) points out that one of the rules for better assessment is that one should always give comments to students about where they could improve, and this was seen as an important aspect of the individual written feedback to students by the markers.

With these new processes in place, the perception shared by the assessment team and myself was that validity of the assessment process related to BP/BHP would improve. With this background in mind, I set about exploring whether in fact this was the case. What was quickly evident was that the question of whether this new assessment process was an improvement on the previous one could not, as I had hoped, be answered with a simple yes or no. Assessment is a complex issue and many factors influence markers in their decision-making processes, irrespective of whether a marker is marking his/her own student or that of another facilitator. I, therefore, became interested in what underlying factors influence the BP/BHP assessors in their decision-making about essay assignment marks, and what are the implications for assessment validity.

## **2.6 Literature review**

Studies related to assessment practices expressed the same concerns about validity, particularly related to the criteria of reliability or fairness of essay assessments, as those experienced in BP/BHP. Linked to these concerns were studies that explored how judgements on the part of markers are made and the potential for bias in interpretations.

Linda Holmes and Lois Smith (2003) investigated complaints from students related to the marking of essays. In seeking clarity, the authors asked students to complete the statement: "It really irritates me when an instructor grades my papers and..." (Holmes and Smith, 2003, p 320). The responses were divided into eleven categories, which could in turn be divided into two broad categories of problems related to fairness in grading and too little feedback from markers. These issues of fairness, specifically about inconsistencies in grading across markers and difficulties with feedback are echoed in the challenges faced in the marking of the BP/BHP essays described above, both from the student and marker perspective.

Alan Branthwaite, Mark Trueman and Terry Berrisford (1981) were concerned with the issue of sources of unreliability in the marking of essay assignments. They asked fifteen university lecturers to mark an essay assignment and to complete an Eysenck Personality Questionnaire. Marks for the essay assignment ranged from 30% to 63%, and a significant positive correlation was found between the marks and the grader's Eysenck Personality Questionnaire for the lie scale. They pointed out that the meaning of the lie score is ambiguous, but is generally an index of social desirability indicating possible effects of staff-student interactions. They concluded that "the need for objectivity and impartiality in assessment can conflict with the aim of establishing positive and co-operative relations in teaching and learning. Depending on the personality of the tutor, considerations of social interaction may bias the marker's objectivity" (p 46). The challenges of impartiality, objectivity and bias in assessment as described by Branthwaite, Trueman and Berrisford resonate strongly with challenges described by the BP/BHP markers who struggled to separate their knowledge about individual students from their performance in essay assignments.

Newstead and Dennis (1994) also looked at the reliability of essay marks by investigating levels of agreement between experienced assessors. They concluded that reliability or consistency in marks was low, and suggested that the number of assessments and assessors should be increased, variety of assessment tasks should be encouraged and that marking schemes should be introduced in order to increase reliability.

Within the South African context there has been limited work on the challenges of essay-type assessments. Reed, Granville, Jank, Makoe, Stein, van Zyl and Samuel (2003) are amongst the few who have explored this area. They investigated what factors produced intermarker reliability and unreliability in the assessment of Honours research reports. They were able to uncover implicit assessment categories, which were shared by markers, but weighted differently. Two issues remained unresolved – that of language and the role of the writer’s voice. Their concluding concern was that consistency across and within universities in the South African context is difficult to achieve, and that a lot actually depends on professional judgement, intellectual position and personal taste.

Shay (2005) conducted a study on final year projects in an Engineering Department at a South African University, which showed that academics sometimes struggled to come to consensus on what marks to allocate. Shay came to the conclusion that assessment is a socially-situated interpretive act (2004) with markers being influenced by a range of subjectivities. (I shall be returning to Shay’s contribution in chapter 3, when I explore her theory of what influences markers.)

Wyatt-Smith and Castleton (2005) conducted a three year Australian study into how teachers make judgements about fifth year student writing. They found that “There is no simple linear course that teachers follow to arrive at their judgements” (p 135), and that teachers used “official” and “other (personal) factors” in their judgements. Wyatt-Smith and Castleton divided these judgements into six indexes that included firstly what they called assumed or actual knowledge about the community in which teachers were working. Teachers would adapt their expectations of students based on the community context. The second area of judgement was based on the teacher’s level of experience in that teachers spoke of this experience as being a resource for their professional judgements. Third was moderation practices as evidenced in markers coming together to share, compare and discuss essays particularly those that they found difficult to mark. Fourth was the assessment criteria and standards set by the markers. Fifth was what they called their observations of students, or what could also be called the marker’s prior and accumulated knowledge about individual students. Finally, knowledge of pedagogy such as those involved in classroom interactions was seen to play a part.

Dennis, Newstead and Wright (1996) conducted a study of the marking of final year psychology undergraduates. They found that if a student was personally known to a marker, marks were “contaminated” by individual biases including generalisations from previous performances and what is called the halo effect, where a marker bases a student’s grade in one assessment on information related to the student’s previous performances. Typically the student who is known to do well will be expected to get a good grade and will be assigned a grade based on this knowledge rather than on his/her actual performance. Physical attractiveness of the student and interpersonal liking were also identified as playing a role in the allocation of marks. Students who were perceived by markers as being physically attractive or more to their liking potentially received higher grades.

Another source of potential bias that was identified was that of group stereotypes where a marker would be influenced by issues such as gender. Newstead and Dennis (1993) looked specifically at issues of sex (or gender) bias in marking of students’ work. They found differences in results for men and women, but no absolute explanation for the differences. They still recommended blind marking although they recognised that even this would not eliminate all biases.

Dennis, Newstead and Wright (1996) posed a model based on their studies whereby they saw a mark as being made up of three components including true merit, marker’s biases, and random influences. They felt this had implications for the reliability of marking.

In 1961, Diederich, French and Carlton developed a rubric for assessment of writing with the aim of making criteria more explicit. Bob Broad (2003) challenged the idea that a rubric or scoring guide is adequate for assessing writing because “traditional rubrics and scoring guides prevent us from telling the truth about what we believe, what we teach, and what we value...” (Broad 2003, p 2). Their weakness is essentially what they leave out. Markers, according to Broad, have to rely on their own prior experience, knowledge and value systems to assist them in interpreting a text. Broad revealed that assessment includes textual (qualities and features of the actual text) as well as contextual (the wider context) criteria. The contextual factors are those which are typically considered irrelevant to the assessment process and kept secret, but which, according to Broad, should be exposed. I shall be returning to Broad’s work in the theory chapter.

What is evident from the above studies is that these writers have also experienced challenges with the assessment of essay assignments, most particularly issues of validity including reliability, and that a number of potential biases on the part of markers were identified. These biases included factors such as where students were known to markers and their relationship came into play (Dennis, Newstead and Wright, 1996 and Branthwaite et al 1981) as well as the marker's experience (Wyatt-Smith and Castleton, 2005) and even personal taste (Reed et al, 2003).

## **2.7 Conclusion**

This thesis explores concerns around the validity of essay assessments, but with the specific goal of exploring what underlying factors influence the BP/BHP markers in their decision-making processes. The above studies will be returned to in the discussion chapter when a comparison will be made between the results of my own study and those of other writers.

University of Cape Town

## **CHAPTER 3: THEORY**

### **3.1 Introduction**

As outlined in the previous Problem chapter, the focus in this thesis is on the underlying factors that influence assessors as they make decisions about what marks to allocate to essay assignments, and the implications these factors and the judgement process may have for the validity of the assessment.

My thesis locates assessment and validity within the philosophical landscape of objectivism, relativism and interpretivism as outlined by Richard Bernstein (1983). Bernstein differentiates between the two major philosophies of objectivism and relativism, and then moves beyond these to outline the interpretive approach, which is the theoretical perspective adopted in this thesis. This philosophical overview will then be followed by an exploration of the potential influences on markers through a framework of concentric circles of both textual and contextual factors (Broad, 2003) within the interpretive approach.

### **3.2 Assessment and validity**

#### **3.2.1 Overview**

As was described in Chapter 1, assessment is a key element of the formal educational process (Ramsden, 1992). According to Gibbs (1999.1), “assessment has six main functions:

1. Capturing student time and attention
2. Generating appropriate student learning activity
3. Providing timely feedback which students pay attention to
4. Helping students to internalize the disciplines standards and notions of quality
5. Generating marks or grades which distinguish between students or which enable pass/ fail decisions to be made
6. Quality assurance: providing evidence for others outside the course to enable them to judge the appropriateness of standards on the course” (p 47).

The fifth function, generating marks, is of particular interest to this study because of its focus on the marking process.

In order to better understand assessment and the different approaches to it, it is useful to first take a step back and explore the wider philosophical perspectives out of which the different approaches to assessment have emerged.

### **3.2.2 Differing philosophical perspectives**

The American philosopher and social theorist, Richard Bernstein (1983) argues that there are broad fundamental questions related to who and what we are, what and how we know, that have been the cause of much debate in both ancient and contemporary philosophy. He describes two opposing poles of philosophical responses to these questions in terms of what he calls objectivism and relativism, and then challenges researchers to go beyond these traditional responses.

Bernstein defines objectivism as “the basic conviction that there is or must be some permanent, a-historical matrix or framework to which we can ultimately appeal in determining the nature of rationality, knowledge, truth, reality, goodness or rightness” (Bernstein, 1983, p 8). Relativism, on the other hand, denies the existence of such claims and argues that whatever philosophers take to be most fundamental, “whether it is the concept of rationality, truth, reality, right, the good, or norms ... must be understood as relative to a specific conceptual scheme, theoretical framework, paradigm, form of life, society or culture” (Bernstein, 1983, p 8). This either/or position is grounded in what Bernstein calls the Cartesian Anxiety. Cartesian philosophy is “the conviction of the universality of reason and the belief that there are universal standards and criteria of rationality” (Bernstein, 1983, p 29). Cartesian Anxiety is the belief that if these standards or criteria of rationality do not exist we will be plunged into chaos.

Bernstein argues for the need to move beyond these two philosophical poles of objectivism and relativism. He draws on the work of philosophers Hans-Georg Gadamer, Jurgen Habermas, Richard Rorty and Hannah Arendt who, although they differ in many ways, share the key themes of dialogue, conversation, communication, communal judgement, and individuals as equal participants in their exploration of the broad fundamental concerns described above. Bernstein also highlights the fact that these philosophers recognise the role of differences, which they say must be resolved through debate and communication. Bernstein sees each as contributing to “exorcising the Cartesian Anxiety and to the

movement beyond objectivism and relativism” (Bernstein, 1983, p 225) to that of an interpretive perspective. Bernstein’s differing philosophical approaches can help us to frame the discussion that follows.

### 3.2.3 Assessment

It is possible to identify various theoretical perspectives or explanations of assessment along what could be seen as a continuum. At one end of the continuum, according to the psychometric tradition, knowledge “must be attained through an objective distance from the world, and if this distance is not maintained, there is a risk of tainting reality with our own subjective beliefs and biases” (Heshusius, 1994 in Angen, 2000, p 380). The primary aim of this approach to assessment is therefore to “control the measurement process in order to ‘isolate the variable’ of pupil performance” (Torrance in Filer, 2002, p 173). This idea of an absolute objective truth resonates strongly with Bernstein’s philosophical perspective of objectivism.

At the other end of the continuum, Wood and Gipps (referenced in Torrance) point out that there has been a shift in understanding assessment away from a focus on measuring individual differences, which is central to the psychometric model, to a more holistic and educationally appropriate one, which privileges learning rather than measurement. This shift has been explained by the fact that “we cannot separate ourselves from what we know, our subjectivity is an integral part of our understanding of ourselves, of others, and of the world around us” (Angen, 2000, p 385). According to Rabinow and Sullivan (1987, in Shay 2004) the goal of this interpretive tradition is to “explicate context”. Context (such as social, cultural, economic and political contexts) is, therefore, all-important, and a neutral and impartial stand, as advocated by the psychometric model, is impossible. The assessor and the assessment process therefore have to be understood within this wider perspective.

According to Moss (1994), in a psychometric approach to assessment each student performance is scored independently by assessors who have no additional knowledge related to the students or about other assessors’ input. In contrast, an interpretive approach emphasises the need to understand the whole in terms of the parts so that the text, the context including the assessors are key. The interpretive approach to assessment would therefore “suggest that because of different perspectives two different assessors will not see the same

thing. (This) creates both intra-marker and inter-marker tensions because they privilege different contextual considerations, different forms of evidence” (Graaff, Reed and Shay, 2004 p 61). Graaff, Reed and Shay point out that different interpretations are inevitable and that a grade is an indication of both “product” and “process” or what can be called the non-cognitive factors in assessment. In other words, a grade is the result of what has been produced by a student as well as the process that the marker goes through in making the judgement.

Shay (2003) draws on Moss (1994, 1996) to identify what really makes the interpretive approach to assessment unique. The emphasis is on the three key areas of difference, context and power. Where the psychometric approach would see difference as a threat, the interpretive approach sees difference as inevitable. Shay points out that difference can, from this standpoint, even be seen as an asset because differences force us to confront what we value, and therefore to confront our own prejudices (Gadamer, 2003, and Bernstein, 1983). Shay also draws attention to context, and through the work of Berlack, Broadfoot and Gipps points out that as Berlack (2000 in Shay, 2003) says “context matters”. Gipps says that “Our mistake, with hindsight, has been to believe that learning can be assessed accurately and reliably... . Not only do we know that performance is context bound, affected by motivation and the assessment task, but more importantly (because it is hidden) performance is construed according to the perspectives and values of the assessor...” (Gipps p 12 in Shay, 2003, p 22). Power is the third critical area identified by Moss. In the interpretive approach, issues of power, such as who makes decisions and the implications of this, are made explicit rather than hidden. In contrast, in the psychometric approach those “most knowledgeable about the context and most affected by the results” are typically silenced (Moss, 1994 in Shay, 2003, p 41).

### **3.2.4 Validity**

Underpinning these different explanations of assessment are fundamentally different understandings of what validity is about. Validity, as was defined in the Problem Chapter, is about the soundness of interpretations. This study is therefore focused on the soundness of the interpretations made by the BP/BHP markers. Maureen Angen differentiates the various approaches to validity by linking them to the psychometric and interpretive approaches described above. Although Angen is looking specifically at research, her discussion of

validity is relevant to assessment. Angen proposes the use of the term validation rather than validity “ to emphasise the way in which a judgement of the trustworthiness or goodness of a piece of research is a continuous process occurring within a community of researchers” (Angen, 2000, p 387).

Within psychometrics, validity and reliability are deemed all-important. Validity is understood to require distance between the reader or marker, because of his or her subjective bias, and the object of study (Angen, 2000). This traditional conception of validity has “evolved from a naturalist view of social science, which ‘maintains that the social sciences should approach the study of social phenomena in the same ways that the natural sciences have approached the study of natural phenomena’”(Martin and McIntyre, 1994, pp. xv-xvi in Moss, 1996, p 21).

According to Moss (1996) complex assessments in which students are required to demonstrate skills of interpretation and the integration of multiple skills and knowledge, have become more prevalent. This shift in assessment approaches has resulted in the need to revisit the more traditional understandings of validity and to recognise the role of professional judgement by assessors. “Interpretive methods can illuminate some of the more subtle ways in which assessment choices interact with the local context, affecting the validity of the intended interpretation and the consequences of assessment use” (Moss, 1996, p 23). “Validity does not need to be about attaining ... objective truth, it lies more in a subjective, human estimation of what it means to have done well” (Angen, 2000, p 392).

Drawing on the work of Bourdieu, Moss and others, Shay (2004) explores the complex interpretive environment of professional judgement, and defines specific contexts for these judgements. “These multiple contexts include: the macro-social conditions of the field and the ways in which these conditions legitimate particular classificatory systems; the meso-level context of disciplinary (and sub-disciplinary) communities of practice with their specific epistemological orientations; and finally the micro-level context of assessor’s interpretive matrices, which are significantly constituted by assessor’s disciplinary interests, professional experience and levels of involvement” (Shay, 2003, p V). She links these multiple contexts to the concepts of “field” or social structure and “habitus” or identity as described by Bourdieu. My research has focused on this micro-level, exploring what factors

influence markers in their decision-making. This micro-level cannot be isolated from the whole picture because interpretive perspectives emphasise that researchers cannot look at the parts without looking at the whole.

### **3.2.5 Potential “threats” to validity**

Crooks and Kane (1996) divide assessment into eight linked stages, considering possible threats to validity at each stage. Of particular relevance to this study are the stages of scoring and evaluation. In the stage of scoring, different markers bring their “own prejudices” to the process that can result in certain people placing more value on certain criteria than others. (This idea of prejudices will be explored in more detail when referring to the work of Gadamer later in this chapter.) This then has implications for validity. Crooks and Kane identify potential threats to validity in terms of scoring. These include the threats that the scoring does not give adequate recognition to all aspects of the task, over-emphasis of some criteria at the expense of others, lack of consistency in marking between markers or by individual markers, and scoring that is either too analytical or too holistic.

Crooks and Kane also identify three threats to validity related to evaluation, the meaning that is attached to the student’s performance. Two are of particular concern to BP/BHP. The first concerns itself with markers who may not have an adequate grasp of the assessment information and may therefore misinterpret the student’s work. The BP/BHP marking criteria are meant to help with this area of difficulty, but as the criteria are not content-specific, this places the responsibility on the facilitators to know the content areas of all the assessments well. The weighting or value attached to different content areas may vary for different facilitators.

The second threat to validity related to evaluation and of concern to BP/BHP is that of the possibility of a biased interpretation of the student’s work by the marker. Nitko (2001), as previously noted, talks about the “halo effect” and the “carryover effect” which essentially mean that students who are known to perform in a certain way are expected to carry this through to other assessments and the marker starts to expect a certain standard from the student, either good or bad.

### 3.3 What potentially influences markers?

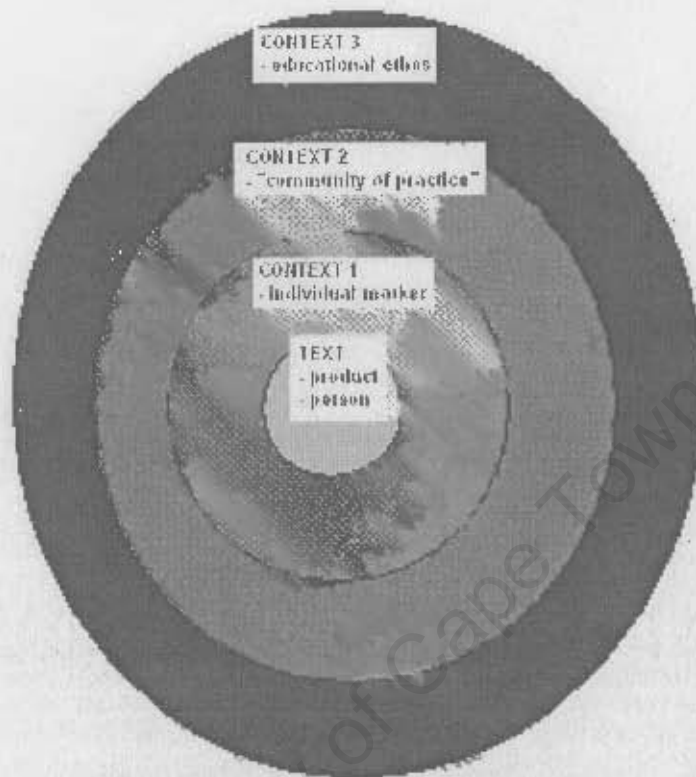
In order to understand what it is that influences markers in making their judgements, it is useful to explore these influences within the hermeneutic circle as described by Hans-Georg Gadamer. Hermeneutics is one particular strand of interpretive theory. Hermeneutics was historically associated with interpretation of Biblical texts, but more recently linked to philosophy, particularly the work of Martin Heidegger and Hans-Georg Gadamer.

Hermeneutics essentially means to interpret. It is a form of understanding through a process of exposing meanings. Richard Bernstein explains that Gadamer equates this process of interpretation with understanding because he claims there is no difference between interpretation and understanding. Furthermore, Gadamer sees understanding, interpretation and application as interrelated because understanding involves interpretation, and interpretation involves application (Bernstein, 1983). For Gadamer, “meaning is always coming into being through the happening of understanding” (Bernstein, 1983, p 139). Therefore it is through the process of interpreting and understanding that meaning becomes evident.

The hermeneutic circle of understanding is “object oriented, in the sense that it directs us to the texts, institutions, practices, or forms of life that we are seeking to understand. It directs us to the sensitive dialectical play between part and whole in the circle of understanding” (Bernstein, 1983, p 135). Textual and contextual factors are therefore critically important, and the whole must be understood in terms of its parts and the parts in terms of the whole.

For purposes of illustration, it is useful to see the potential influences on markers as four concentric circles or parts (see diagram) – with textual and contextual factors being explicitly delineated (Broad, 2003). The inner circle includes the textual factors of the person, or student, as well as the product, which refers to the essay itself. The second circle is the first of three contextual factors, the first referring to the individual marker, and the next being about the individual within the team or community of practice. The outer circle refers to the pedagogical ethos and context of the course itself within which the team or community and individual markers function. It is important to note that these four “circles” are part of one another and these parts are what make up the whole; understanding one part is not sufficient – all parts or contexts need to be understood.

## Factors that influence markers



### 3.3.1 The inner circle: Textual factors

The innermost circle includes the textual factors (Broad, 2003) that may potentially influence markers. Textual factors refer to both the product, the text or essay itself, as well as the person or student, the writer of the essay.

### 3.3.2 Contextual factor 1: The individual marker

The first contextual area of potential influence (Broad, 2003), but the second circle, is what I have described as the individual marker. Central to Gadamer's theory is the role of pre-judgements or prejudices, in other words what the interpreter brings with him or herself to the text. "Gadamer stresses that we must always temper our understanding to the 'things themselves'; we must listen to them and open ourselves so that they can 'speak to us'; we must be receptive to the claims to truth that they make upon us. But on the other hand, we do not do this by bracketing or forgetting all our pre-judgements and prejudices. On the contrary,

it is only because of the play of these prejudgements that we are enabled to understand the 'things themselves'" (Bernstein, 1983, p 138). This focus on the role of what are variously referred to as prejudices, prejudgements, fore-structures or fore-knowledge is central to Gadamer's work and critical for this thesis. Gadamer sees a person who is trying to understand a text as always projecting because the person reads with certain meanings or prejudgements which he or she then attaches to the text as a whole as soon as some initial meaning becomes evident. Understanding and interpretation involve a constant process of projections, which are replaced by new ones as new information emerges. What is interesting is that Gadamer does not see prejudices as only negative. For him a prejudice is "a judgement that is rendered before all the elements that determine a situation have been fully examined" (Gadamer, 2003, p 270). Although prejudices are typically understood in a negative light, they can, according to Gadamer, be productive when they facilitate understanding and therefore have a positive value. "Prejudgements and prejudices have a threefold temporal character: they are handed down to us through tradition; they are constitutive of what we are now (and are in the process of becoming); and they are anticipatory – always open to future testing and transformation" (Bernstein, 1983, p 140 – 141).

According to Moss, hermeneutic philosophy "recognises that the reader's preconceptions, enabling prejudices, or foreknowledge, are inevitable and valuable in interpreting a text. They are in fact what make understanding possible. Here, the hermeneutic circle is viewed as including a dialectic between reader and text to develop practically relevant knowledge" (Moss, 1994, p 7).

The process of the hermenutic circle is, according to Gadamer, a constant process because new meanings are always emerging in a movement of understanding, interpretation and meaning. This "process of understanding can never achieve finality... We are always understanding and interpreting in light of our anticipatory prejudgements and prejudices, which are themselves changing in the course of history" (Bernstein, 1983, p 139). Understanding, for Gadamer, is therefore always evolving as people go through and are exposed to new and different experiences. "The important thing is to be aware of one's own bias, so that the text can present itself in all its otherness and thus assert its own truth against one's own fore-meanings.... It is the tyranny of hidden prejudices that makes us deaf"

(Gadamer, p 269 and 270). It is through the hermeneutic circle or process that one can reach understanding of the conditions and assumptions underlying one's own views, prejudices or prejudgements. Interpretation, according to hermeneutics, therefore occurs through time in a circular and iterative way. From a hermeneutical perspective, practical dilemmas, for example different interpretations based on different values, are inevitable. It is through the process of the hermeneutic circle that one is able to reach resolution within a particular context and within the constraints of reality such as time.

### **3.3.3 Contextual factor 2: The “community of practice”**

Gadamer sees the hermeneutic circle as the way in which a community of readers engages with a text. The second context of the whole or third circle is, therefore, the team of markers and the process they go through. In order to elaborate on this context, the insights of situated learning are useful to draw on, in particular the work of Etienne Wenger and how communities of practice operate. According to Wenger, the individual functions in the context of the team or community of practice. Integrative judgements are reached through extensive discussion, and “a final decision represents a consensus or compromise based on that discussion” (Moss, 1994, p 8).

According to Wenger, a person's identity is made up of an interplay between individual and community aspects. Identity is viewed in social terms without denying individuality. “It is therefore a mistaken dichotomy to wonder whether the unit of analysis should be the community or the person. The focus must be on the process of their mutual constitution” (Wenger, 1998, p 146). This understanding of community and its influence on markers and vice versa resonates with the interpretive focus of the interplay between part and whole. It is therefore useful in extending our understanding beyond the individual into the team or community of markers.

When Wenger speaks about community, he is not referring to a geographical community, but rather a grouping united by practice. According to Wenger it is practice that helps to make sense of the community itself. “The concept of practice connotes doing, but not just doing in and of itself. It is doing in a historical and social context that gives structure and meaning to what we do. In this sense, practice is always social practice” (Wenger, 1998, p 47). This idea of practice as social links strongly with the interpretive view of the practice of

assessment as a social activity. For Wenger practice provides help in resolving institutionally generated problems, encourages a communal memory so that individuals can focus on their own tasks without having to know the entire picture, and assists newcomers to join the community through the activity of the practice itself. It develops frameworks for the achievement of specific tasks, and also helps to create an environment conducive to achieving despite some of the more tedious tasks that may be involved (Wenger, 1998 p 46). These concepts of community, identity and practice are useful in explaining how individuals come together with their prior experiences and knowledge and form a team to work on a specific task such as the marking of essays.

Communities of practice exist through what Wenger calls mutual engagement, joint enterprise and shared repertoire. In terms of mutual engagement, individuals, who often are not a homogenous group, come together and organise themselves around a common purpose. Diversity is inevitable, but it is the relationships that are created between people that, despite differences, ensure the community is sustained. It is important that individuals feel and experience being included in the community's activities. Complementary and overlapping roles must be acknowledged, and tensions and conflicts must be managed through a process of "community maintenance" where underlying dynamics are given recognition.

Mutual engagement happens through what Wenger calls joint enterprise in which individuals go through the process of negotiating meaning for the community and ultimately developing mutual accountability amongst community members. The broader explicit and implicit contexts from which individuals have emerged have to be taken into account, as well as the fact that any joint enterprise is not static. This obviously has implications for power because all aspects that define the relationships between members have to be negotiated and then renegotiated through time.

Finally, community membership and coherence facilitate the development of a shared repertoire in terms of a discourse particular to that community. An example of this would include health professionals who develop a language or jargon particular and essential to their functioning together as a community of practice. In terms of assessment this shared repertoire would develop through the use and language of marking criteria, training sessions and moderation sessions.

Wenger points out that a community of practice can be a positive site of strength or a negative site of weakness, a source of creativity or a source of oppression. So much depends on the factors of mutual engagement, joint enterprise and shared repertoire.

Gadamer also looks at these factors, but in terms of prejudices. He describes what he calls the horizon as everything that can be seen from a particular vantage point. Through language and shared contact with the horizons of others, Gadamer believes communities are forced to confront personal prejudices and resolve differences.

### **3.3.4 Contextual factor 3: The educational ethos**

What has been identified here as the outer circle includes the broad context of what can be called the educational ethos. This includes the programme within which the assessment is taking place as well as the department, faculty and wider university context.

Within the broad philosophy of interpretive theory and more specifically hermeneutics, Paul Ricoeur identified two opposing positions – that of suspicion, which is drawn primarily from the work of Habermas, and that of faith, from the work of Gadamer. Ricoeur explained suspicion as a more critical, judgmental perspective, reflected in the school of critical theory and typical of Freud and Marx. He defined the perspective of faith as a more empathic and sensitive approach, grounded within hermeneutics. Graaff, Reed and Shay, 2004, drew on Ricoeur's distinctions and related these opposing views to how judgements are made. Both views recognise assessment as a form of judgement, but from different perspectives within an interpretive framework. What is significant about this is that this overarching value system becomes the wider pedagogical ethos for the assessment activity. Whether a more critical approach typical of a purely summative assessment, or a more supportive approach, such as in-course assessments, is advocated and evident, it will influence how assessors see their role and engage in the task of judging a student's work.

### **3.4 Why an interpretive frame?**

This thesis is framed within a broadly hermeneutical and interpretive framework because I believe, like Shay (2004), that assessment is a socially situated interpretive act. By this, she means that "assessment practices can only be understood within the social, cultural, and

economic contexts in which they occur..., and the experiences, beliefs and expectations of both the assessed and the assessor constitute the meanings of assessment outcomes” (Shay, 2003, p 1). This is particularly appropriate here in the context of BP/BHP because essay assignments are complex tasks, which reflect the “integration of multiple skills and knowledge, and require expert judgement” (Moss, 1992, p 230).

In understanding assessment from an interpretive approach, it is “the *context* of assessment, the inevitable *difference* in assessment interpretations, as well as the equally inevitable effects of *power*” (Shay, 2003, p 16) that are most significant. As Knight (2002) points out, the practice of assessment is really about judgement rather than measurement because where measurement is exact, judgement and therefore assessment are not. Difference from a psychometric point of view is recognised as a problem with reliability, which in turn affects the validity of the assessment. From an interpretive perspective, however, difference is recognised as inevitable and potentially useful as a resource that can help to highlight the varying prejudices or prejudgements that different markers bring to the experience. The interpretive perspective therefore encourages recognition, openness and discussion of these differences. Framing this study within an interpretive perspective means that differences in factors that influence markers must be explored.

## **CHAPTER 4: METHODOLOGY**

### **4.1 Introduction**

As explained earlier, this thesis began with a particular interest in whether my decision, as course convenor, to not allow facilitator markers to mark their own students' work was the correct decision. The process of collecting data led me to the more complex question of what underlying factors influence markers' decisions.

The data for this thesis was collected from six sources using mainly qualitative methods. The qualitative methods included:

- Individual interviews with BP/BHP facilitators who were also members of the marking team
- A video recording of a moderation session
- Student feedback on the assessment process
- An audio-taped moderation session
- Marker written reflections

In addition a comparative study using quantitative methods of analysis was conducted.

### **4.2 Validity of the study**

In terms of validity of the research process, Fraenkel and Wallen (2000) refer to "appropriateness, meaningfulness, and usefulness of the inferences researchers make based on the data they collect" (2000, p 506). This resonates with Nitko's emphasis on the soundness of the interpretations. Validity in this instance was strengthened by the fact that a range of sources was used. This allowed for a process of triangulation whereby different sources and methods are used to reach conclusions. The process of documentation, audio-recording and video-recording meant that accuracy of the source data and analyses was increased.

### **4.3 Individual interviews with facilitator markers**

I began my research by interviewing ten BP/BHP facilitators who were part of the BP/BHP marker team in 2004. Interviewing is described by Fetterman (in Fraenkel and Wallen, 2000) as the most important type of qualitative research because it is how researchers can check

their impressions and assumptions for accuracy. It is the process of finding out from people those things that cannot be directly observed (Patton in Fraenkel and Wallen, 2000, p509). Qualitative research interviews are professional conversations “where the outcome is a co-production of the interviewer and the subject” (Kvale, 1996, p xvii). In my study, I was in the dual role of both interviewer and course convenor. As convenor I work very closely with the BP/BHP small group facilitators, who are the subject of this study. I select, train and support the facilitators who in turn are answerable to me. In this convenor-facilitator relationship, the establishment of trust is central and I do this by building empathic relationships with my facilitator team. This had both advantages and disadvantages in the interview process. The closeness meant that the markers already knew and trusted me, and were therefore more likely to be prepared to speak openly about their experiences and opinions. The disadvantage of our relationship may have been the desire on the part of the markers to say what they thought I would most like to hear, and to hold back their criticisms and opinions because of their desire to support me in my study.

I chose this method of data collection because I was looking for in-depth accounts of the marking experience that would allow me to gain insight into the markers’ point of view. I chose semi-structured rather than pre-structured questions because the questions needed to allow for open-ended responses and space for the further exploration of answers and themes as they emerged.

The interviewees consisted of ten out of a total of twelve markers. My intention had been to interview all twelve markers, but time did not allow for this. The interviewee group consisted of nine women and one man. Ages ranged from early twenties to mid-fifties. Experience as facilitators on BP and BHP ranged from a few months to three years. All markers held tertiary qualifications, ranging from Social Development (Social Work) to Law, Education, Occupational Therapy and Psychology. Despite the diversity of the group, what was shared was an interest in and commitment to the BP/BHP course outcomes as well as the assessment process.

Each interviewee was asked to read and sign a letter of consent agreeing that the contents of the interview could be used for the purposes of this study, and that their real identities would not be revealed (see Appendix 4).

The interviews took place between the months of August and November 2004, and were each of thirty to forty-five minutes duration. They took place in my office in the Faculty of Health Sciences at the University of Cape Town as this was a convenient venue for both the interviewees and me. I made use of semi-structured questionnaires (see Appendix 5) with the aim of exploring whether the interviewees thought the new marking system was preferable or not and why. I audio-taped each of the interviews. After completing the interviews, each interview was fully transcribed and then I began the process of analysing the data in order to establish themes.

Fraenkel and Wallen (2000) suggest that semi-structured interviews are most useful at the end of a study as they facilitate what can be called the drilling down into particular hypotheses. In the case of this thesis, it may have been premature to conduct the interviews at this early stage in the process particularly as time did not allow for re-interviewing.

#### **4.4 Video recording of a moderation session**

As the individual interviews, described above, progressed, it became apparent that further insights into the marking process would be gained by observing markers in action. This would also provide an opportunity for triangulation of my findings. I therefore asked the marking group's permission to observe and video-tape an assessment moderation session in which the markers come together to moderate their marks and comments to students. The group, which consisted of the team described above, agreed to me observing and video-taping their moderation session on 27 August 2004. My motivation was a desire to see markers in action dealing with the types of challenges they face when "defending" their marks and comments to other markers. I video-recorded the session in order to review the session afterwards.

This method of research placed me in the role of observer rather than interviewer, enabling me to observe the markers as they actually went through the marking and moderation process. Video-recording the session enabled me to identify theme areas and particular critical incidents of interest, which then informed, challenged and supported the findings from the interview process.

#### **4.5 Student feedback**

Having now begun to get a picture of what markers thought about the BP/BHP marking process in the individual interviews and what influenced their marks in the moderation session, my research turned to what students thought about the new assessment process, and what concerns they might have.

At the end of 2004, all students who participated in the BP/BHP courses were required to complete a detailed evaluation form, which included two questions related to the new marking process. Students were asked to respond to the following statement by indicating their responses on a scale of strongly disagree, disagree, agree or strongly agree. The statement read: "I like the fact that my essay assignments were marked and moderated by facilitators who were not my group facilitators" (Course Evaluation, 2004). Students were then asked to comment on their responses. Three hundred and thirty-four students completed the evaluations in BP/BHP group time. The data was collected on-line in the students' information technology (IT) classes. This was an efficient process that meant that the number of responses was high and the data could be analysed almost immediately.

#### **4.6 A comparative study of differences in marks**

The data to this point had been collected from interviews with markers, a moderation session and feedback from students. One of the themes that emerged from the interviews was the perception that knowing the student would have an impact on the mark assigned to a particular student. I decided to set up a study to measure whether there was in fact any statistically significant difference between marks assigned by markers who marked their own students' work as opposed to the marks assigned by a second marker who did not know the students.

Students were required to hand in their first essay assignment on Friday 18 March 2005. They were informed that they were required to hand in two copies of their assignment as two markers would be assessing each essay independently. Two hundred and twenty-four of the three hundred and forty-eight essays were marked by the students' own facilitators because that was the number of students whose facilitators were also markers.

Markers were divided into marking pairs by the assessment co-ordinator. A preparatory marking training session was held on the morning of Friday 18 March 2005, and I addressed the group, explaining that I wanted each person to mark their own students' work and those of their partner. I gave each marker a handout explaining the process (see Appendix 6).

That afternoon, after all assignments had been collected, they were given out to the markers. Each marker in the marking pair received papers belonging to his/her own student group and papers from their partner's group. They were specifically asked not to compare marks until the moderation session on 1 April 2005.

On Wednesday 30 March 2005, markers sent their final individual marks to me. On Thursday 31 March, I entered and compared all the marks. I then sent the marks to Mr Rauf Sayed, a Senior Biostatistician in the School of Public Health and Family Medicine at the University of Cape Town, to analyse whether in fact there were differences. Using the Bland-Altman plot, markers who assigned significantly different marks were identified (Kirkwood and Sterne, 2003).

Markers came together on Friday 1 April 2005 for a moderation session in which they were asked to compare and resolve marks.

#### **4.7 An audio-taped moderation session**

Having received permission to audio-tape and observe markers in moderation after the comparative study, I joined the group in moderation on the afternoon of 1 April 2005. Once again, I used observation to gain further understanding of what and how markers come to consensus about marks. I had four tape recorders with me, one of which proved to be damaged. As a result three out of seven pairs were recorded in moderation, while the other four pairs were observed. The pairs I chose to tape were Lea and Janet, Nicole and Tanya, Anna and Tamsin, as they represented a cross-section in terms of professional discipline, experience in facilitating BP/BHP and experience with marking.

After I listening to the audio-tapes from the three pairs as they compared marks and worked together to come to an agreement as to what marks and comments would be most accurate, I transcribed all discussions of relevance.

#### **4.8 Facilitator-marker written reflections**

After the marking and moderation process, I sent out an e-mail to all members of the 2005 marking team inviting them to comment on the marking process. Seven responded, and I compiled these comments into a word document, and analysed the responses in terms of the themes already identified in the previous interviews and moderation sessions.

#### **4.9. Data Analysis**

Having collected data from these six sources, and having spent time analysing each portion, I then began analysing the data as a whole, seeing it as interactional rather than distinctly separate. Themes began to emerge, and these formed the basis of the data that is analysed in the following chapter.

The main advantage of my study methodology was that of triangulation, comparing and checking findings from different sources. The main weakness of these study methods was that I was forced to interview and observe those markers who were available and I was unable to re-interview interviewees because of time constraints. It is possible that some of the findings may not have accurately reflected what the interviewee or marker being observed or interviewed said. Furthermore, I did not include non-markers in my study, and it is possible that they might have added different insights.

#### **4.10 My role as researcher**

My role as researcher in this study has been two-fold. On the one hand I am convenor of Becoming a Professional and Becoming a Health Professional, and therefore have a vested interest in the outcome of this study in terms of the decisions I have taken and will take in the future. I also have insights from years of being involved with the courses. On the other hand, I am interested in and committed to sound educational practice and therefore want to analyse the assessment practices in BP/BHP with this in mind.

Throughout the process, I felt the pull between qualitative and quantitative data. Working in a Health Sciences Faculty, where scientific evidence of a quantitative nature is typically more valued, I found myself looking for absolute answers. It was not long before I realised that this drive for a “yes” or “no” response would not be possible because I was looking at

essentially social processes. An interpretive view was required so that the richness evident in the qualitative data, which takes into account the context of the study, could be fully explored.

University of Cape Town

## CHAPTER 5: DATA ANALYSIS

### 5.1 Introduction

As outlined in the Problem chapter, this study began with the explicit intent of finding out if I had made the right decision, as course convenor of BP/BHP, in no longer allowing the small group facilitators to mark their own students' essay assignments. My interest in this question was motivated by the need to verify whether this new assessment process had addressed the concerns related to validity of the BP/BHP assessment process. I was particularly interested in concerns around "subjectivity" and possible bias on the part of those markers who were grading their own students' essay assignments. Collecting and analysing the data made it clear that a simple "yes" or "no" in response to my initial question of whether I had made the right decision in no longer allowing markers to grade their own students' essays was not going to be either possible or sufficient. There was a much deeper question of what influenced markers that needed to be explored in terms of the validity of the assessment process. Markers and students alike had a very particular understanding of what they associated with validity – in their words validity, or the soundness of their interpretations, was about fairness and objectivity which they associated with distance between the student and marker. The new system should then have allayed their concerns, but what in fact emerged was a much messier picture.

This chapter presents the data collected and analysed while exploring this thesis topic. As validity is the central concern throughout, the chapter begins by revisiting this concept from the view of markers and students before delving into the messier picture of what factors influence markers in their decision-making processes.

### 5.2 What is validity?

Although the meaning of validity was explored in the problem chapter, it is important to recap exactly what is meant by validity in the context of this study because it is such a central concept. For my purposes, my research drew on Nitko's definition, which describes validity as "the soundness of your interpretations and uses of students' assessment results" (2001, p 36). According to Nitko, reliability, a concept which is often identified as something separate from validity, is one of the criteria for or strands of validity and is specifically concerned with consistency in marking and fairness in grading. The key element

related to validity that is pertinent to this research is therefore the soundness of the markers' interpretations, specifically consistency across markers who may have different perspectives. For students and markers alike this consistency is equated with fairness.

### **5.2.1. How do the BP/BHP markers understand validity?**

Validity for markers needs to be understood in the context of markers wanting to do justice to both the students and the course. According to the data, they seemed to feel that they were answerable for their decisions to the students, to me as convenor and representative of the course and to the wider Faculty and University contexts. They knew that I had changed the marking process from the initial system of facilitators marking their own students to a system whereby markers could not mark their own students' work in order to encourage increased fairness and consistency in marks. They also felt the pressure from students, the majority of whom supported the process of markers not marking their own students' work on the grounds that they perceived this as ensuring more fairness and less bias.

Although they never explicitly referred to the term validity, it is clear from the data that the BP/BHP markers believed that they were best able to make sound, or what they referred to as "fair", interpretations if they were not marking their own students' work. This was most clearly evident in that all the markers interviewed supported the change from marking their own students to only marking students from other facilitator-lead groups.

Chloe said: *"I am very pro the fact that I don't mark my own students' essays. Because I am not aware of where the students' academic abilities lie, or their written abilities lie, I treat everyone the same in the group. We all try to be objective, but there is a limit, and you always are influenced by, you know, is it a male or a female, is it a student that is nice to you...."*

Karen said: *"I completely support the decision that we don't mark our own students... I think the more familiar we are with the students, the harder it is to be objective."*

Ella supported the new system of marking, but expressed ambivalence: *"On the other hand I feel that I have a more mentoring kind of relationship with them about the development of their writing skills when I am marking my own group."*

Markers, for the most part, therefore equated the ability to be fair or make sound interpretations with distance between student and marker.

Despite my belief and that of the markers I interviewed that knowing the students would influence their marks, in the comparative study of marks given by their own and another marker, of the fourteen marking team members, the marks of only two markers were significantly different (see Appendix 8 groups 4, 5, 11. See also 12, which was borderline). Nicole and Gail, whether they were grading their own students or students they did not know, gave the essays higher marks. This may be because it was only Nicole's second semester on the marking team, whereas it was Gail's first time marking with the BP/BHP team. Overall, whether markers were marking their own students or the essays of students they did not know, the differences were not significant (see Appendix 7). Looking at that particular data, it would seem, therefore, that whether a person marks their own or another student's work concerns about bias and fairness related to markers knowing students were, according to this data, unfounded. But, on closer examination the qualitative data told a very different story from the start. This data, taken from both students and markers, indicated that there were clear areas of concern with regards to markers marking their own students' work. What also became clear to me was that marking is a complex process and that whether a marker is marking his/her own student or that of another facilitator, many factors come into play.

In elaborating on this need to be distanced from the students in order to make sound decisions, markers spoke about the struggles they experienced. The most clearly identified areas of challenge for markers when marking their own students were

- where a student was identified as someone the facilitator particularly liked, or
- where a student was perceived to be struggling, but had shown evidence of making an effort in group.

The issue of favouritism with a struggling student came through most strongly in one of the audio-taped moderation sessions between Lea and Janet when they were discussing a discrepancy in the marks for one of Lea's students. Despite assessment training and explicit

marking criteria, Lea struggled between her affection for the student and his actual performance which seems to have been poor.

Janet: *"This one has also failed"*

Lea: *"I can't. I love him. He's my favourite."*

Janet: *"Squeeze him through."*

Lea: *"I can't. He hasn't answered the question. He is just so sweet, but... He hasn't referenced at all. He's misunderstood the question. I don't know if he should pass. He's a mature student and so insightful, but I know he can't pass."*

Janet: *"Let's pass him onto Alfred (for moderation)."*

In this extract, what is clear is that Lea is struggling intensely to separate her own experience of the student based on her contact with him in her group from his actual assessment performance in his essay assignment. Despite her struggle, Lea seems about to come to the conclusion that the student cannot pass the assessment based on his performance in the task, when Janet takes the responsibility away from Lea by suggesting that they leave this difficult decision up to someone else.

Markers reported struggling with their desire to mark the students' performance as evidenced in their assessment task against their desire to reward and encourage students they perceived as struggling, but making an effort or trying hard. They therefore reported finding it very difficult to separate the students' performance in group from what he or she had produced for the essay assignment. This was evidenced in the previous interaction between Janet and Lea as well as in the following discussion.

In the audio-recorded moderation session, Janet and Lea struggled with marking a second language English student, a member of Janet's group, privileging her with particular knowledge related to the student:

Janet: *"This is definitely a student who has a problem with English as a first language. She is from a rural area in the Eastern Cape."*

Lea: *"I gave her 54 for that section"*

Janet: *"No, I gave her 70%."*

Lea: *"That's because you know she has a problem and is a second language English student. ...There was no true understanding. This could be because she doesn't understand the question."*

Janet: *"So, you don't think she can even get 60%?"*

Lea: *"If we give her 60%, she is not going to see where she went wrong."*

Janet: *"She has a huge problem expressing herself because she is from a very rural area of the Eastern Cape. So, I think it's there, but she doesn't know how to get it out. But that is also because I know her...."*

Lea elaborated, in her interview with me, about this difficulty of knowing students:

*"You know when you can't put a face to a name in front of you it is easier and you become more objective, and you read that piece of paper for what it is, as opposed to linking it to the person.... You have this perception OK, well they are dedicated and they work extra hard. And when you read the paper you think, you know, should you reward them for that, so you do become subjective there...."*

Here Lea speaks about the importance of objectivity in the assessment process, but also questions whether in fact one should value the subjective knowledge of a student in some way. She calls the marking of the assessment "easier" when she does not know the student because then she perceives her interpretation to be based purely on the written task work, which for her means that the interpretation is objective and therefore sound.

In the markers' desire to be objective and distanced from their students, markers revealed that they put measures in place to avoid knowing the identity of the student they were grading. Anna said, in her interview with me, that she tried never to look at the name of the student on the cover of the essay because she felt sure that she would be influenced by her own knowledge and experience of the student. She described it as: *"It's a bit like opening a Christmas present, you know... if you have clues you will go to the clues before you open the present. But I try to not do that. Because it immediately puts in your mind "oh, this is a 70% student or that's a 60% student."* For Anna, again this idea of distance between marker and student was what she perceived to be important for fairness.

I experienced my own inability to remain either quiet or objective during the video-taped moderation session debate about the appropriate mark for one particular student. This student had been to see me earlier that day because she was having a very hard time coping after an extremely traumatic event. She had been very tearful and overwhelmed by her drop in grades and the possible effect this would have on her future in the Faculty and on her

personal relationship with her parents. With this information fresh in my mind, I found myself interrupting the discussion and asking the markers to “*Go with the higher mark – trust me on this.*” Within moments, I apologised and asked the moderating pair to please ignore my interruption and allocate the mark according to the criteria. I had not been able to separate my knowledge of the student from her performance in this particular essay assignment – in fact I had no knowledge of what her performance was as I had not even read her essay. I was concerned about the issue of fairness if the mark was pushed up when perhaps there were other students as deserving of our sympathy.

In that same video-recorded moderation session, Jean relied on information about two students that she knew through another course. She defended the actions of the students in the following quotations: “*I think that was a mistake. She is a very conscientious student and then: Knowing X, I think she excluded the in-text referencing in her word-count.*” In an interview with me, Jean had been absolutely certain that there needed to be objectivity and distance in the marking process, and yet in action she struggled with this herself.

### **5.2.2. How do the BP/BHP students understand validity?**

As with the markers, the BP/BHP students also never referred explicitly to validity, but inferences can be drawn from their responses to the course evaluation questions posed at the end of 2004. Again, “fairness” was of greatest concern. The majority of students supported the process of markers not marking their own students’ work on the grounds that they felt it assured more fairness and less potential for bias on the part of the marker.

When students were asked to respond to the evaluation statement “I like the fact that my essay assignments were marked and moderated by facilitators who were not my group facilitators” with strongly disagree, disagree, agree or strongly agree, eighty-one percent of the three hundred and thirty-four students responded that they liked the fact that their assignments were marked by another facilitator-marker, whereas nineteen percent indicated that they did not.

Asked to comment on their answers, students elaborated on why they felt positive or negative about their essays being marked by a facilitator who was not their group facilitator. Of the two hundred and fourteen (sixty-four percent) comments that supported marking by a

facilitator other than their own, the most frequently used words were “fair” or “unbiased”. These specific words were found in one hundred and thirty-four of the responses. Comments from students included:

*“I like to receive an opinion about my work from someone I do not know. I feel that since they have no preconceived ideas about me, they would not show bias when they mark my essay.”*

*“It makes one more comfortable as one knows that the marking process is fair. You needn’t be scared that the facilitator has a bad impression of you from the group and that that may spill over into the marking.”*

*“It is better and fair and professional that way. The moderator doesn’t know me and so there cannot be any bias or favouritism.”*

Another word that was linked to fairness and evident in twenty-two (ten percent of the responses) was “objective.” A student commented: *“Because the (marker)-facilitator does not know me, they will mark my assignment objectively.”*

What is interesting, however, is that of the nineteen percent of students who at the end of 2004 indicated that they would have preferred to be marked by their own facilitator, the vast majority indicated that their facilitator knowing them and their abilities or style would have made for more informed marks and feedback. As one student said: *“(Your own) facilitator knows you much better as a person and therefore may understand where you are coming from and therefore understand your logic behind a specific argument, opinion or answer.”*

This evidence therefore pointed to the fact that students perceived the relationship between facilitator and student as potentially influencing the allocation of marks. One student elaborated:

*“Sometimes you find that a facilitator is very close to a student. In that situation, you cannot guarantee that the mark is exactly the mark you deserve.”*

A few students commented on the subjectivity of the whole marking experience whether it was their own or another facilitator. Examples include:

*“Does not really make a substantial difference (who marks my work) because it is still highly subjective to the marker and his/her mindset.”*

*“I feel that although a level of bias may come into play in the marking of essays, one must realise that a level of subjectivity is found with any marker marking any essay.”*

To summarise before moving on, markers and students supported the changes in the marking because they perceived it as a more fair and objective process. They were concerned about bringing biases related to specific students into the assessment. It was this knowledge and experience of individual students that markers perceived as potentially effecting the soundness of their interpretations and their ability to be fair and consistent in their marking. The evidence that emerged from the data, however, revealed a still wider range of factors that could be seen to potentially jeopardise fairness on the part of markers. As Gipps (referenced in Torrance, 2000) points out, assessment is about social beliefs and values. Assessment reflects the value judgements of powerful groups, in this case that of the markers and the assessment co-ordinators. These can also be understood as individual prejudices or prejudgements (Gadamer, 2003) that markers bring to the task of marking. These factors, or value judgements, shall now be explored in more detail in order to highlight the complexity of validity and the “soundness of our interpretations” beyond actually knowing the student him or herself. These areas of influence are divided into four categories, under the two main categories of textual and contextual factors (Broad, 2003) as identified in the theory chapter:

- The inner circle of textual factors
- Contextual factor 1: the individual marker
- Contextual factor 2: the “community of practice”
- Contextual factor 3: the outer circle - the educational ethos

### **5.3. The inner circle: Textual factors**

The inner circle is made up of textual factors referring to the written product, that is the essay as well as the person or writer of the text, in this case the BP/BHP student.

Although textual factors are what is evidenced in the written work itself, it is important to remember that one cannot understand the part without looking to the whole. Boundaries between what is described as textual or contextual are not always clear, and in many cases should be seen as shifting.

Markers said in their interviews that the most important factor in the marking process was the essay or text itself, and identified quality as the key influence in their allocation of marks. Certain aspects stood out as most important to the markers in the video-recorded and audio-recorded moderation sessions when markers discussed and negotiated the final marks for each essay assignment. These included the following areas:

- ***Understanding and clarity of the knowledge***

In moderation, this verifying of understanding and clarity of knowledge was most evident when markers looked to see whether key concepts had been defined and explained. They consistently checked for accuracy by looking at the criteria and referring to Jean, the assessment co-ordinator, if they were unsure.

- ***The weighting of content versus structure and language***

Many markers struggled with how to weigh content versus structure and use of language in allocating marks. There seemed to be agreement throughout that a well-structured essay using clear correct English was evidence of a good essay because of its positive impact on the reader, but just how good and how much value to attach to this was the question that markers struggled with most frequently.

As Lea said in her interview with me: *“It is important, it makes such a huge difference if you read an essay which just has a continuous flow. It makes quite a big impact on the reader, because it makes you understand it more.”*

Two aspects stood out. Firstly, some markers felt that just by reading the introductory paragraphs they would already have clues about the quality of the rest of the essay. They linked this to their own experience of having marked essays of what they described as good as opposed to poor quality work. For them a good essay typically began with a clear, well-phrased introduction.

Secondly, the markers that described themselves as being poor at English because it was their second language or because they struggled with learning challenges admitted to assigning less importance to structure and language than other facilitator-markers. As Anna said in her interview with me: *“I still struggle with the whole thing of language, you know.”*

*How much should the student's ability to write English properly effect the mark... Because I'm a second language speaker I have sympathy for that and empathy. Ja, you can see what he or she is trying to say doesn't exactly come out that way, and then you know it's quite difficult to allocate a mark to a paper like that. Because how much do you penalise a person for lack of good, you know, writing skills or English language skills."*

- ***Academic writing***

The word "academic" came through repeatedly in the moderation sessions, particularly when markers grappled with what they thought was a suitable style of writing. What was not clear was what they actually meant by "academic" – it would be thrown into the conversation as a means of justifying a mark particularly where a facilitator-marker was defending a lower mark. Janet tried to explain what she was looking for in the audio-taped moderation session when she was pushing for a student's mark to be lowered: *"Her writing is too colloquial. We need to have academic writing...She was too emotive for an academic essay."* It would seem therefore that what Janet equated with academic writing was a more formal style where emotional, more personal insights would be inappropriate.

- ***Correct and accurate referencing***

In their interviews, personal reflections and audio-taped moderation session, referencing was highlighted as being important. Tanya summed this up: *"And I think it is also important for the students to know the referencing properly, because if you are going to write, you know, any academic work and your referencing isn't correct, it creates like a very bad impression."*

In the moderation sessions, markers discussed how much they should be allocating towards accurate referencing. An example of this was when Lea and Janet, who had agreed on 65% for an essay, began to question their mark when they looked at the referencing.

*Lea: "But would you give anyone 65% for not referencing, without referencing at all actually?"*

*Janet: "For a first essay I would, but maybe I am doing it wrong."*

*Lea: "Maybe I am being too strict."*

*Janet: "I just think you are being a little hard on him."*

*Lea: "Let's go with 65%, but word the comment about the referencing quite strongly."*

- ***Following instructions***

For each essay assignment in BP/BHP, instructions as to layout, structure, etc are very clear (see Appendix 6). The word limit is specified for each essay and is explained to students as being important because when they graduate as health professionals and begin to write for journals they will need to write both accurately and concisely. It is a challenged area because frequently markers ask students to explain or elaborate on a point causing the students to ask how this can be done within the given word-limit. This was also a challenge for markers, most evident in the video-recorded moderation session when markers finally agreed that the only way to be sure that students were reflecting the word count accurately by counting the words. This decision was based strongly on the need to be fair particularly to those students who had worked within the word limit. As Janet said: *“I think it’s hard on students who have stuck to the word limit if we don’t take marks off those who have gone over.”*

- ***Effort***

Markers spoke in the moderation sessions about the evidence of effort in the student writing regardless of the fact that they did not know the student they were marking. They seemed to want to attach a value to effort, but how it could be identified was unclear. This was particularly the case in essays that were borderline pass-fail where markers were reluctant to fail a student. In the video-taped moderation session, Tanya struggled with this issue as she worked towards finalising a mark: *“When I got this essay, it felt very messy.... I don’t know if it is just me, it just feels as if there wasn’t that much effort put in. It’s probably a subjective thing to say, but presentation does count. To me it says something about the effort the student has put into the work.”*

#### **5.4 Contextual factor: 1– the individual marker**

This second circle is concerned with the individual markers and the factors identified as influencing them.

- ***Educational and professional background of markers***

The BP/BHP markers are, as described in the Problem and Methodology Chapters, from very different educational and professional backgrounds. Many saw these background areas as influencing them and were aware of the importance of clarifying their role in this particular setting, for example not marking in the role of English teacher, lawyer, psychologist etc. From the interviews in particular, markers seemed to find their

backgrounds to be both an advantage and a disadvantage. Ella, who was an English teacher for many years said: *“I am very critical of English being used properly, so I have had to be very careful of that. And when you are used to marking English as a subject, then you know, you’re obviously critical. Occasionally I will see something very clear, like an apostrophe being used incorrectly and I will actually write a little note in the margin saying the apostrophe, you are using is an apostrophe when it’s just a plural. I will feel, oh Ella you are going back into English teacher mode. But sometimes I think, you know, it often isn’t just taught and maybe that person is going to use that, and they are not going to confuse the issue later on. But if I can see that generally they are really battling, sometimes I’ll just, I’ll add in a word so that the sentence makes sense, or something like that. But I try not to do too much of that.”*

Rene, who is a lawyer by profession said: *“My background makes me quite rigorous ...that is probably an influence of law, you know be careful of the words that you use and choose because they have different meanings. I have been trapped within my own discipline, and it’s taken an effort for me to get out of that. So there is that tension.”*

Two of the markers spoke in their interviews about having been brought up in a home where English was not their first language, and how this influenced them. In the audio-taped moderation session, Anna explained to Tamsin that she struggled to mark one of her students because of her own issues with language: *“So, this is one essay that I think I marked really badly. Sometimes she says things and because I am Afrikaans speaking, I can relate to that and know what she is trying to say when it comes out as something totally different.”*

- **Marker Experience**

Various markers commented on the value of having prior marking experience, and how these influenced their interpretations. Karen, Tanya and Lea had all been involved in marking Psychology assignments. Karen commented that: *“Marking lots of psychology work has certainly helped”*, but added that having recently been a student herself she could recall what sort of comments and feedback had been useful. She could now draw on this experience in her role as marker.

Nicole was one of the new markers who really struggled and who stood out as giving marks significantly different to those of her marking partner. She said: *“I will mark and then doubt myself and go back to it and mark them over. I really enjoy the process of moderation. That to me is where I counteract my insecurity, to have someone else who is more experienced looking at my marks.”*

Anna, who has been a facilitator and marker on BP since the inception of the course commented: *“I am much more confident than I was three years ago, in terms of marking these particular essays. Because I know the content better, I’m a better marker now.”*

Alfred, who was part of the BP team for the third year added *“I definitely bring experience now, and am able to support some of my newer colleagues.”*

The markers who had been involved in the design of the assessments indicated that participating in that process had been of great value. As overall assessment convenor, Jean said: *“Being involved, so involved in developing model answers and formulating that grid, means that I think I am sort of more prepared than most of the facilitators, who are still digesting that kind of information.”*

Chloe who was part of the core assessment design team said: *“Being on the marking team with Jean and Anna, where I’m very clear on what is expected, what should be included, and I am very clear on the concepts...I have that framework in my head and it really helps.”*

- ***Other personal influences***

The markers identified a range of other more personal influences that they felt impacted on their marking.

Karen spoke extensively in her interview about how her own academic drive influenced her marking: *“And I think I have always been quite a marks orientated person. Getting a first was the most important thing and you know getting 60’s is just not on... And ja, I suppose I feel like, that it’s only fair to the students that we put the effort in and that we can justify (their marks)...”*

Karen also spoke about how her own enjoyment of writing had influenced her *“I really enjoy writing. I enjoy pulling things together, I enjoy that really creative process when you have to start from nothing and produce something. If it’s something that you enjoy or that you feel passionate about... then I think that being involved in that process of helping students can be really good.”*

Alfred said in his interview, and wrote in his reflective feedback, that his own personal challenges affected his ability to mark: *“I am dyslexic. As a result thereof I can quite easily be forced to read a paper two or three times before I am able to fully grasp what the student is trying to say. I guess a result of dyslexia is a difficulty in reading.”* What was interesting for me was that Alfred saw this as a challenge but also a positive influence in that he felt he had real empathy for students who struggled.

The age and experience range is also an influence. The facilitator team members have always ranged from new graduates to those who are more experienced and mature. Each end of the age spectrum has particular advantages and disadvantages. Tanya, who is one of the younger team members and a new graduate commented on the advantages of being a marker at this stage of her own development: *“I have been a student quite recently, so it’s quite fresh in my mind how I’m supposed to write and sort of what the academic style is.”* Jean, a more mature member of the team, commented *“I have experience on my side, but I do worry that my age is a disadvantage when working with such young adults.”*

Particular opinions about the student and their performance were evident whether markers actually knew the students whose papers they were marking or not, based on what they felt the essay said about the student. An example of this was related to where markers struggled over papers written in what they perceived to be the student’s second language. Markers seemed to be able to pick up from the writing that the student was a second (or third) language English speaker. Various dilemmas came into play. This is obviously of particular interest in our South African context because it is so linked to our history. This was best summarised by Rene’s input in her interview with me: *“I gave this person 50, but the English was very bad. He had the basic understanding, but just couldn’t get it across... I don’t like to hold that against people, because I mean that I think it was a Black student or someone who was writing in a second language, who has not had the same benefit of the*

*education system that I have had. But yet understands the stuff, but can't actually express it. But it is very difficult because there is that side of you, and then there is the other side that thinks that well we are all supposed to be professionals and (students) are supposed to become professionals. And then their ability to write will be looked at and criticised by people who won't bear in mind those sorts of extenuating circumstances. I mean it is a very difficult tension there. So in that way my sort of political feelings and my general feelings about SA and its past influence how I mark, I think."*

Another area of potential influence, which could be inferred from the student's name, is that of gender. International evidence that looks at the factors that influence markers sees gender as a very significant area of influence (Newstead and Dennis, 1993). In this study, it was not an issue for many markers. It was, however, not a specific question. Rene was the exception in that she saw this as a potential area of concern because by knowing the sex of the student she said she would begin to get a mental picture of the student in her head.

Evidence of negative unsubstantiated views or prejudices expressed by students was not an area that came through as influencing many markers, but Chloe expressed her concerns so strongly in her interview with me that it is important to include this information here. Chloe: *"I find that I can be particularly judgmental on certain scripts where students have written something that goes against kind of my grit. Is that the right word? Um, and you know it's a funny thing I'll suddenly find my handwriting is getting bigger across the page, because I am actually angry with this student for making sweeping comments or very judgmental statements. And I find I have to then take a step back from that and be more aware of my own reactions and responses to it... One essay spoke about the fact that the people in the community were unemployed and it's because they just don't have the wherewithal to think of ways on how to make money. And they should get off their bums.... I battle not to get angry at them. I battle sometimes not to put a big fat pencil mark through the page. I actually have to stop myself from doing that."*

## **5.5 Contextual factor 2: the "community of practice"**

The above factors can be understood as areas of personal or individual prejudice that influenced markers. It was, however, through the shared marking process that individual markers spoke of or showed evidence of confidence in their final judgements. The interface

between individual and community activity therefore became the context for negotiating meaning (Wenger, 1998).

Jean, as assessment co-ordinator, described the moderation process, in our interview. She explained the process:

- Markers are paired and after grading their own allocation of scripts, they exchange scripts with their co-marker. The co-marker does not know what mark the initial marker has assigned. The idea is that *“every essay is marked and moderated by a second marker by the time people come to the final moderation session.”*
- Markers meet for a final formal moderation session. These final, formal moderation sessions last three hours in total.
- The first hour of the moderation session is spent in pairs as they finalise the marks through a process of discussion of similarities and discrepancies.
- The full marking team is then divided into sub-groups for the next hour to ninety minutes. One subgroup is responsible for checking essays with grades of seventy-five percent and above (first class passes), another for seventy to seventy-four percent (upper seconds), another for grades in the sixties (lower seconds), another for fifty-three to fifty-nine percent (third class passes) and finally another for fifty-two percent and below (borderline and fails). The group that is responsible for fifty-two percent and below looks at every essay in that range, so each will have been seen three times. For all other grades, each marker pulls out one script in each range and gives them to the moderation team. That way a cross-section of all essays is moderated for consistency across markers.
- Finally, all marks are entered on the departmental lists.

In their interviews, a number of markers spoke about the responsibility of taking the marking seriously and being *“committed to the assessment process”* supporting the rest of the marking team and the process as a whole. Ella referred to this as *“loyalty to the team”*. This resonates with what Wenger calls mutual engagement through supportive interpersonal relationships and the activity of community maintenance.

Both markers and students referred to the value of the moderation process, or what could be understood as the joint enterprise as described by Wenger, in supporting the allocation of appropriate marks. Students commented that they felt reassured knowing that their essays marks were not assigned by one person, but went through a process of checks and balances.

For markers, the value seemed to be linked to the support markers felt in the moderation sessions. This was both of a personal nature because markers could share their marking difficulties and learn from each experience, and a professional nature because they saw moderation as a chance to double check their work and have the team as a whole take responsibility for the marks. They, therefore, valued not feeling alone in their decisions.

In her interview with me, Chloe summed up the value of moderation: *“The moderation process... is the best thing. I find that in terms of the students’ marks, we do really discuss them and people feel comfortable enough to say “Hey, I don’t agree with this”. And that makes a huge difference. So I find it’s very, very beneficial when discrepancies are picked up and then sorted out.”*

Tanya also commented extensively on the moderation process: *“I like the fact that you mark it and you get someone else to moderate, and you sit for a whole afternoon, it really takes out a lot of that subjectivity, you know. And people do, like even if you’ve got marking criteria, everyone does mark differently. So to actually sit with someone else and to go through it and to consolidate the way you mark is really helpful. You know if I was a student I would be very grateful that it was done like that. It helps immensely. Because sometimes you don’t notice things in an essay that someone else might pick up on. So I think that is really good.”*

In addition to the moderation process, the marking criteria were greatly valued by the markers. This was reiterated in the individual interviews, reflections and moderation sessions. Recording the two moderation sessions enabled me to observe markers “in action” as they went about resolving what marks to allocate. Markers placed the criteria for the essay on the table in front of them. They checked their marks and comments against the criteria and used the criteria to justify a mark and to form the basis of written feedback for students. In her interview Chloe commented: *“I keep the criteria with me, always, and then I go back, and I also try and make the comments against the criteria.”*

For many markers there was still concern about the actual grades they should assign because of the wide range of marks within each criteria category. As Rene said: *“Generally the marking table is quite helpful, but then there is still ten percent leeway, and that really is a*

*gut feel for me, I must be honest. I have an idea of what an average essay is like, and then I kind of work from that... I think of what they say to the student. So for me a mark of like fifty-nine will say a lot to a student compared to a mark of like fifty-seven, because a fifty-nine is almost a sixty and it is like I am purposefully saying that it is not a sixty because you just, you don't make it."*

Evidenced above is what Wenger calls the development of a shared repertoire whereby the discourse around BP/BHP marking add meaning to the process.

Time constraints were identified by markers as a negative influence in the marking process because of the pressure to be finished within an assigned time-frame and the essentially repetitive nature of marking. This resulted in marker fatigue. Ella summed this aspect up: *"I struggle with the boredom of reading the same stuff over and over. I think it's the sheer hours - time and of fatigue."* Chloe linked this to the challenge of maintaining the same standard throughout because as she put it: *"After thirty essays everything becomes a blur."* This was supported by many other facilitator-markers who said they struggled with how to keep interested and consistent in the marking process after reading similar answers one after the other. Nicole found it was the last five that she found extremely difficult to mark: *"I find it difficult to stay objective as I am very tired at that stage. I feel like I have read it all."* Alfred reflected on the quality of his own marking: *"I find by the time I reach the last five papers I become exceptionally critical of smaller things and foolish errors that I would have been more tolerant of when I started marking."*

### **5.6 Contextual factor 3: The educational ethos**

This marking of students' work happened within the context of the BP/BHP courses with their explicit outcomes and methodology. Markers were drawn from this context as they were all first and foremost facilitators of student groups. The methodology of BP/BHP is largely experiential and a significant level of trust is therefore encouraged within groups. While the groups are educational in focus, there exists at times a fine line between the educational and the therapeutic because of the way in which the groups function. This links with Ricoeur's explanation of faith and suspicion as described in Chapter 3. Facilitators who became markers experience a significant shift in their roles from supportive and non-

judgemental to judgemental and critical. Both roles are part of the educational ethos of the courses, but can be challenging for markers to hold together .

In support of the new marking process and linking this to the educational ethos of BP/BHP, Rene said: *“All in all I didn't like having anything to do with marking my group's papers because I think that my objectivity was compromised. This is due to the nature of the sessions and that we build up a strong bond with each student.”*

### **5.7 Difference, context and power**

As was described in the theory chapter, the areas of difference, context and power are central to an interpretive approach. Context is described in detail through the concentric inner, middle and outer circles. Also found in the study data were significant differences in interpretations by markers. This was most evident in the moderation sessions when markers presented their grades to other markers, went through the process of discussing the differences and similarities, and came to a grade. Power in terms of who made decisions in the marking process was evident as markers turned to Jean and myself for final decisions. But what was also clear was that the markers felt confident in the process that had been set up because they frequently turned to other team members in times of difficulty.

### **5.8 Summary**

The BP/BHP courses provided the context for this study of what influences markers in their decision-making processes and the implication for validity. Evidence from the data collected for this study pointed to the fact that markers and students alike equate validity or the soundness of interpretations with fairness and objectivity. Markers therefore perceived distance from the students to be of importance for making sound judgements. Despite this, a range of textual and contextual factors such as the essay and writer, the individual marker, the marking community and still wider issues of the educational ethos were evident as areas of influence.

## CHAPTER 6: DISCUSSION

### 6.1 Introduction

To recap, this thesis came about as a result of my work as course convenor of two first year Health Science courses called Becoming a Professional and Becoming a Health Professional (BP/BHP) in the School of Public Health and Family Medicine at the University of Cape Town. As course convenor, I became involved in a range of educational issues, including assessment.

In 2002 and 2003, the small group facilitators of BP/BHP were responsible for marking their own students' assessments including essay assignments. As was described in Chapter 2, a number of concerns arose related to the validity or soundness of markers' interpretations when grading their students' essay assignments. Of particular interest to this study was the concern that markers of their own students' work might be biased in favour of their own students because of their knowledge about and relationship with the student. Linked to this concern was the fact that facilitators found it difficult to shift from the more supportive role of small group facilitator to that of a more judgmental role of marker. Students in turn raised concerns about the fairness or accuracy of marks assigned to them.

In response to these concerns, particularly that of possible bias if markers knew the students they were grading, a new system of marking the BP/BHP essay assignments was introduced. From 2004 the BP/BHP small group facilitators were no longer required to mark essays. A sub-group of facilitators with a particular interest in assessment was formed under the leadership of an assessment co-ordinator. In order to overcome the problem of possibly jeopardising the validity of the assessment process, markers were no longer allowed to mark their own students' work. This information was shared openly with the students.

In 2004, I began research into whether it had been the right decision to no longer allow small group facilitators of the BP/BHP courses to mark their own students' essay assignments. I was looking for a clear answer, but the answer, as evidenced in the data presented in Chapter 5, was not straightforward in that it was not a simple yes or no. A more complex issue came to the fore in the form of the deeper question of what underlying factors influence markers in

their decision-making about essay assignments and what are the implications for assessment validity?

## **6.2 The data**

The data from the quantitative comparative study in which essays were double marked, first by the student's own facilitator and then a second marker, revealed that overall there were no significant differences between marks assigned to students by their own group facilitators and markers who were not their facilitators. This was unexpected as the changes that had been instituted had been largely in response to the concern about the influence of bias from markers who knew the students. The reason that there were no significant differences may well have been a result of the extensive preparation in terms of training for marking, and the provision of marking criteria. The training allowed for clarification of tasks as well as open and clear discussion while the marking criteria meant that all markers had a clear reference point. Had these not been part of the process, it is possible that the results for the quantitative comparative study may have revealed something different.

While the quantitative analysis indicated no discrepancy between the marking by the students' own facilitators and those of the non-facilitators, the qualitative data collected through interviews, observations and reflections revealed something quite different. All the markers that were interviewed perceived themselves able to be more fair if they were not marking their own students. They expressed greatest difficulty with marking the work of students from their own groups if they identified these students as either struggling, but trying hard, or as what they called "favourite". This was also supported by the vast majority of students who, in their end of course evaluations, said they preferred that someone other than their own group facilitator mark their essays because of concerns around bias and fairness.

What became clear from the data was that the factors that influence markers and the soundness of their interpretations are many and varied; from what the text itself contains to marker experience, from the marking criteria and moderation sessions to the broad context of the educational ethos of the course. There is no easy technical solution to strengthening validity because it is a process central to which is a commitment to opportunities for dialogue and debate around differences of opinion and perspective.

### 6.3 Factors that influence markers

Using Broad's categories of text and context, the factors that emerged as influencing markers in their decision-making processes were presented as a diagram of concentric circles. As was described in Chapter 3, it is important to remember that these divisions within and between circles are fluid in that the parts need to be seen in terms of the whole and the whole in terms of the parts.

The inner-most circle of the diagram included the textual factors of the written product itself and the person/ student or writer. Markers identified the actual quality of the essay as being most important in the marking process. They included demonstration of understanding and clarity of content knowledge as key. They struggled however with different views on the weighting of content versus structure and language, as well as the importance of academic writing, correct and accurate referencing, and the demonstration of an ability to follow instructions. Whether markers knew the students or not, they struggled with a desire to reward effort as demonstrated by the student through the text. Contextual factors as represented by the next three concentric circles of potential influence included the context of the individual marker, which moves then to the individual as part of the marking team or what Wenger calls the community of practice. The overriding educational ethos or context of the course then forms the outer all-encompassing circle within which the individual and team must function. All of these contextual factors impact on marker's interpretation of the text itself.

From the data it was clear that whether markers assess their own students or not, they come with their individual prejudices (Gadamer) which means that they value things differently. Individual markers brought their own prejudices to the BP/BHP experience as they grappled with what marks to assign. As Gadamer argues: "The prejudices of the individual, far more than his judgements, constitute the historical reality of his being" (Gadamer, 2003, p 276). These individual prejudices are linked to the circle of the individual areas, namely the educational and professional background of markers, marker experience and what was broadly referred to as other personal influences. As Rene commented in her interview: "*We are the products of ourselves.*" In other words, each person is a unique mix of prejudices and prejudgements based on their prior life experience. These more personal prejudices or areas

of influence were also supported in the literature, for example, Reed et al (2003) conclude that professional judgement, intellectual position and personal taste influence markers in their decision-making. Wyatt-Smith and Castleton (2005) found that teachers used “official” and “other (personal) factors” in their judgements and that experience was one key influence.

The data also revealed that the next contextual factor of the marking team or community of practice, as a critical area of influence in the assessment process, involves individuals coming together to form a team or what Wenger describes as a community of practice. In the BP/BHP marking process it is the interface between these individual and community practices that ultimately determines the outcome or assigned mark for each individual essay, and supports the premise that assessment is a social practice. As Bernstein points out, the community must be open to “choice, deliberation, interpretation, judicious weighting and application of universal criteria, and even rational disagreement...” (Bernstein, 1983, p 172). In the case of BP/BHP this was most clearly evident in the formal moderation sessions where markers were encouraged to argue and deliberate through the process of coming to what they as a team perceived to be a fair mark.

The outer circle of the educational ethos is the broad frame within which the whole marking process takes place. Markers spoke about the struggle within BP/BHP of moving between the context of the small group where the facilitator-student relationship is valued and nurtured to the task of assessing where distance, objectivity and judgement in the name of fairness are all-important. This can be linked to the work of Paul Ricoeur, in which he identified the two opposing positions within hermeneutics – that of faith and suspicion (Ricoeur in Graaff, Reed and Shay, 2004). In the case of BP/BHP, the student-facilitator relationship, the course content and methodology could be perceived as a perspective of faith because the learning environment is student-centred and very supportive. Relationships between students and the small group facilitators are often strong as they go on a journey of learning together. Facilitators are specifically chosen because of their interest in small group, experiential education with young adults. Course content is mentally and emotionally challenging in that the students are required to develop as knowledgeable, empathic and reflective health professionals. The weekly group sessions challenge students to dig deep in terms of their learning. The focus in facilitator training is not only on the content for group

sessions but also on understanding group process. Facilitators are therefore encouraged to be empathic and sensitive to the students' learning processes.

The BP/BHP assessment process could, in contrast to the group sessions, be understood from a perspective of suspicion. Markers are required to step back from their empathic and supportive facilitator roles to become critical and judgmental as they mark. Explicit pre-determined, objective criteria become the measures for performance. The text itself is taken out of the wider context of the student performance in relation to group activities and group process, and what was valued previously in group is not taken into account in the marking process.

The challenges that emerged in the BP/BHP assessment process may well have to do with a struggle between these basic perspectives of faith and suspicion. Facilitators are required to move from a perspective of support and non-judgmentalism in their small groups to that of critical judgement as markers. This would certainly explain the struggle facilitators experienced with marking their own students, but also that of the shifting role generally – from facilitator to assessor.

#### **6.4 Validity**

The central concern throughout this thesis has been that of validity, the soundness of markers' interpretations or what markers and students understand as issues of fairness and objectivity on the part of markers. Consistent with the literature (Holmes and Smith, 2003) the BP/BHP markers supported the belief that validity can be achieved through distance between markers and students.

Evidence from interviews, observations and reflections suggested that the greatest struggle for BP/BHP markers was with grading students known to them – especially those they perceived to be favourite or struggling students. This concern is supported by Alan Branthwaite, Mark Trueman and Terry Berrisford (1981) who found that social interaction between markers and students may bias a marker's objectivity. Dennis, Newstead and Wright (1996) also found that if a student was personally known to a marker, marks were "contaminated" by individual biases. As was pointed out earlier, however, the comparative study in which marks assigned by the students' own facilitators and those assigned by a

marker not known to the student were not significantly different. This result was therefore at odds with the other methods of data collection used in this study as well as studies cited earlier.

Crooks and Kane (1996) identified specific threats to validity at each stage of the assessment process. These threats, which were outlined in Chapter 3, such as the overemphasis of certain criteria at the expense of others, were echoed in the data of the present study. Crooks and Kane see these factors as threats, with the implication being that validity cannot be assured if these threats or factors are in place. An interpretive perspective sees these factors as inevitable and potentially positive, and draws attention to context, which should be acknowledged and valued as differences are inevitable. The challenge is how to do this while ensuring validity by making sure that the markers' interpretations are sound.

There is no easy technical solution to strengthening validity because it is a process and central to the process is a commitment to opportunities for dialogue and debate, around the differences of perspective which emerge. From an interpretive perspective, differences are accepted as inevitable. It is therefore the marking process that needs to be scrutinised as the place where differences can be resolved. What follows are suggestions for resolving differences and increasing validity in the marking process by drawing on the diagram of factors that influence markers.

Beginning with the inner circle of the text itself where the product as well as the student him or herself are focused on, it would seem useful to explore the merits of marking one's own students as opposed to having someone who does not know the student doing the marking. Newstead (1996) and Lockett and Sutherland (2000), amongst others, support the idea of anonymous or blind marking to reduce bias by markers. What is evident from the present study is that although the marks themselves showed no significant differences, the markers and students were supportive of markers not marking their own students. This needs to be taken into account when weighing up the potential pros and cons of adopting anonymous or blind marking.

Marking criteria, also suggested by Newstead (1996) and Lockett and Sutherland (2000), can be part of the marking process so that all markers have common points of reference. But as

Graaff, Reed and Shay commented, "Even when marking criteria are made explicit, they cannot do justice to the complexity of the interpretive frameworks which assessors are implicitly drawing on." (2004, p52). Since all writing is open to interpretation, the second step should be that all markers participate in a training or preparation for marking session.

In terms of the contextual factor of the individual marker, Luckett and Sutherland (2000), like Gadamer, suggest that markers should know their own values and prejudices and that, like in the context of the community of practice, internal moderation be used to reduce inconsistencies. Differences in understanding of the criteria can be discussed and resolved, therefore giving recognition to the individual prejudices that markers bring to the process. Once the markers have gone through the task of marking their assigned essay scripts, a useful step may be for them to exchange papers with a second marker to begin the process of moderation. This should then be followed by a formal moderation session in which differences between first and second marks are discussed and resolved within the supportive context of the marking team. A final step in the moderation process could be sorting the papers according to grades and then comparing them for consistency across the range of markers. This could also provide the opportunity for discussion of what actually constitutes a certain grade according to the criteria. Issues of difference can in this way be brought into the open and discussed in a supportive environment. Criteria should also be available to students. Meetings between students and the marking team should facilitate open discussion with use of the criteria. Every participant in the process can give input and make changes in consultation with other markers. Finally, the educational ethos needs to be taken into account in terms of what the course is trying to achieve within the Faculty and discipline. In the case of BP/BHP, the aim is to begin to develop students into Integrated Health Professionals within their disciplines of study. This aim needs to be explicitly explained to and understood by markers so that the educational ethos of the course can be truly valued.

### **6.5 Strengths and shortcomings of this study**

Turning now to the strengths as well as the potential shortcomings of the present study, the courses BP/BHP, in which the study took place, form a very specific context. The courses have specific aims and outcomes that may not be linked to or appropriate for other contexts. The method of assessment, however, is widely used and therefore the findings could be useful for other contexts that make use of essay assignments.

The diagram of factors that influence markers, which was designed specifically to illustrate the findings of this study, may be useful as a tool for clarity, but is inevitably reductionist in that the parts appear as distinct rather than open to movement between and across the boundaries.

My own dual roles as course convenor and researcher of BP/BHP could be seen as an advantage or a disadvantage because of my vested interest in the outcome of the study. As course convenor I am the person responsible for the courses – their design, implementation, assessment and evaluation. I am also responsible for the appointment of facilitators. My relationship with the markers meant that I was able to access information that they may not have wanted to share. On the other hand, they knew what I was studying and may have wanted to support me by giving me answers that they assumed would be most appropriate.

In terms of the methodology, quantitative and qualitative methods were used and this allowed for triangulation of the results. Audio and video-recording the interviews and moderation sessions meant that material could be reviewed as and when it was appropriate. In terms of increasing the validity of the interview process, a more in-depth study could include follow-up interviews to check the accuracy of the data. The video recording was useful as a record, but had limitations in that only certain aspects could be focused on at any one time. The audio tapes were effective, but limited by resources.

A more in-depth study could include further input from the students as sources of clarification with regards to their responses in the evaluation questionnaire.

## 6.6 Conclusion

Analysing the data for this study within an interpretive perspective has helped to highlight the complexities of the marking process. It has also provided room for the interpretation of the data within a frame of understanding the overall picture in terms of its parts. Issues of power, difference and particularly context have been highlighted.

The act of marking is about judging and being judged and therefore the BP/BHP markers and students valued objectivity and distance between marker and student as a measure of validity, the soundness of the interpretation. The pull, therefore, always exists towards a desire for absolute answers, typical of objectivism and the idea of a “permanent, a-historical matrix or framework” as described by Bernstein (1983, p8). In reality, however, essay marking is a socially situated interpretive act, and it is the processes that are put in place that ensure validity. The influence of the varying textual and contextual factors must be taken into account.

University of Cape Town

## CHAPTER 7: CONCLUSION

### 7.1 Overview

This study came about as a result of concern about the validity of the assessment process, specifically with regard to essay assignments, within two first year Health Sciences courses at the University of Cape Town. A change was instituted in response to concerns about bias and subjectivity on the part of markers, and markers were no longer allowed to mark their own students' essay assignments. The question as to whether this was a good decision formed the basis of this study, but could not easily be answered. A study of marks assigned by markers who knew the students as compared with those assigned by another marker revealed that it made no difference whether students were marked by their own small group facilitator or another marker. This was a surprising find in light of the data collected through interviews and observations. What began as a seemingly simple question evolved into a study of what factors influence markers in their decision-making processes, and the implications for validity.

Evidence that assessment is truly a complex socially situated interpretive act has been one of the key findings of this study. Interviews, observations and reflections told a different story to the comparative study in that markers and students alike believed the new system to be more fair and free of bias. They equated this with distance between the student and marker. Digging more deeply, specific factors emerged that influenced markers. These were represented in a diagram of factors that influence markers, and included the recognition of markers as individuals with their own prior knowledge, experiences and values, or what Gadamer refers to as prejudices, as well as markers as members of a team or community of practice (Wenger). These roles were then framed within an outer circle of the educational ethos of the courses themselves. Central to all these contextual influences was the text in terms of the essay itself and the person who wrote it.

### 7.2 Implications for Practice

This study has added support to the voices that have recognised assessment as a socially situated interpretive act. What has become clear is that issues of power, difference and

particularly context that are central to an interpretive perspective, must be recognised and explored if the assessment process is to be taken seriously.

Within this study, issue of difference and context were highlighted. Differences in interpretation on the part of markers were evident. Those involved in assessment need to recognise the complexity of the marking process and to include checks and balances to enhance the validity of the assessment process so that differences are recognised, explored and resolved. In terms of context, the diagram of factors that influence markers may prove useful as a tool for others involved in the planning, implementation and evaluation of assessments. It provides focus areas, but these may need to be changed to suit different assessment contexts. What is important to remember is that whatever the contexts are, they must be understood as fluid in that the whole needs to be understood in terms of its parts. The divisions are there for ease of reference rather than as separate entities. Despite the limitations of the diagram, it provides a theoretical lens for understanding the assessment process. In this way, it paves the way for further debate and exploration of what influences markers in their decision-making and what the implications are for validity.

## Appendix 1

### **Becoming a Professional and Becoming a Health Professional 2005**

#### COURSE OUTCOMES

BY THE END OF THE YEAR STUDENTS WILL:

#### **1. BP specific outcomes**

- 1.1. have experience and a working knowledge of how groups evolve and function
- 1.2. have a working knowledge of how people interact and what facilitates good interpersonal skills between individuals and in groups
- 1.3 be able to demonstrate basic interpersonal and interviewing techniques

#### **1. BHP specific outcomes**

- 2.1. have a basic understanding of Primary Health Care- its origins, and philosophy and implementation in practice
- 2.2. begin to value the contribution of different health professionals in the promotion, maintenance and support of health and health care of individuals, families and communities
- 2.3. have basic knowledge and experience of qualitative research
- 2.4. have practical experience in applying the above knowledge, skills and values within a community oriented project

#### **2. Generic BP and BHP outcomes**

- 3.1. be aware of the importance of professionalism in their interactions with colleagues, clients and the public
- 3.2. reflect an understanding of and respect for diversity/difference
- 3.3. have a basic knowledge of the concepts of Human Rights and Health Rights and their implications for practice as a health professional
- 3.4. have developed basic information literacy (IL), information technology (IT) and academic literacy (AL) skills

## Appendix 2

### BECOMING A PROFESSIONAL 2005 – essay assignment

Read through the following case scenario, and follow the instructions below.

#### Scenario

Sister Samuel is a 43 year old Clinic Sister working at Sidney Street Clinic, close to the University.

Cindy is a shy 18 year old first year University student, who is away from home for the first time. She rents a small room in a nearby suburb from an elderly couple.

Cindy tends to get stomach cramps and feel nauseous when she is anxious. On the morning of Cindy's first class test, she wakes up feeling very ill and decides to visit the Sidney Street Clinic.

The day before Cindy visits the Clinic, Sister Samuel finds out that her husband of 20 years is leaving her for a younger woman who is pregnant by him.

The following excerpt is taken from the consultation between Cindy and Sister Samuel.

- Sister Samuel: Yes, can I help you?  
Cindy: Uhm, uhm...  
Sister Samuel: Yes, what is it?  
Cindy: [*fidgeting with her bag*] Please Sister. My tummy is sore and I feel sick.  
Sister Samuel: How long have you felt like this?  
Cindy: Oh, it started in the night, but I often feel like this when I'm stressed.  
Sister Samuel: Do you have a boyfriend?  
Cindy: Sorry?  
Sister Samuel: Could you be pregnant?  
Cindy: No! I, I, I,.....  
Sister Samuel: Come on now. That is what you young girls do these days, isn't it?  
Cindy: No, no, I don't have a boyfriend. I hardly have any friends here! I'm actually really lonely, and we have a test today and I am... [*Sister Samuel gets up abruptly, causing Cindy to stop talking*]  
Sister Samuel: Let me examine your stomach. Do you think you have eaten something strange?  
Cindy: I often get like this when [*Sister Samuel sits down again*]  
Sister Samuel: I can find absolutely nothing wrong with you. It is all in your head. You need to pull your self together and be thankful that you are not really sick.  
Cindy: [*starts to cry*]  
Sister Samuel: What is wrong now? [*looking surprised*]

#### Assignment questions

1. Drawing on
  - class activities,
  - your reflections (using your personal journals), and
  - the chapter on Knowing ourselves from Gibson et al (2002)

Identify how Sister Samuel's professional conduct and empathy were affected by her personal circumstances in this scenario and suggest how self-reflection may be helpful in her dealings with patients in the future.

In your answer explain the concepts of self-reflection and empathy.

Suggest ways in which Sr. Samuel could build up a capacity for self-reflection.

(800 to 900 words - 50 marks)

2. Drawing on your
  - understanding of the integrated health professional
  - your reflections and
  - the reading by Taylor et al (1993)

Describe Sister Samuel's approach to health in this scenario.

Explain what you think are the shortcomings of this approach to health for Cindy.  
Identify which approach to health may have been more helpful to Cindy and why.

(800 to 900 words - 50 marks)

**TOTAL MARKS: 100**

**Structure and Presentation:**

- **Your role as a writer**

Don't think of your role as that of a student writing for a facilitator. You are a professional communicating your insights about the topic to an audience who do not form part of your BP group. You therefore need to clarify what you mean by all relevant terms and concepts. You are being introduced to professional writing skills which are part of your overall training as a health professional as you will one day participate in conferences and publish in your field. (You also need to reflect on your own thoughts, feelings and behaviour when answering the questions)

- **Academic language and style**

You may write in the first person.

Do not use any abbreviations and contractions (like etc, e.g. don't).

Do not use colloquialisms and informal language (like "I freaked out", "That was really cool").

Avoid using bullet points.

Make use of relevant subheadings.

Acknowledge the sources of your information in the text at all times, and at the end of your essay, and include a full bibliography at the end of your essay.

- **Introduction**

Here you give the reader an overview of your answer. This includes a summary of the argument of your answer.

- **Body**

The main point of each paragraph should link coherently and logically to the main points in the previous and subsequent paragraphs (use linking words and signposts such as "therefore", "in the first place", "on the other hand").

You may want to use headings to distinguish different sections of your answer.

- **Conclusion**

No new information must be included at this point.

Round off your argument.

If you want to make recommendations or highlight particular issues, do so at this stage.

**Points to remember**

- Your answers must be typed using size 12 font and double spacing
- The BP assignment cover page must be attached and signed
- Your assignment must not exceed the word limit and must include a word count.
- All students are expected to use the Harvard method of referencing.

**Students handing in essays late:** less 5% for the first day, and 2% for everyday thereafter.

**Essays that exceed the word limit:**

150-200 words over the limit – less 2%,  
200-500 words over the word limit – less 5% and  
more than 500 words over the limit – less 10%

**Problems and queries:**

Members of the marking team will be available to meet with you individually before each assignment is due. Please check the notice board for times and venues.

Appendix 3: BP ASSESSMENT CRITERIA

<b>FAIL Inappropriate</b> 0 – 39%	<b>FAIL Poor</b> 40 – 49%	<b>3<sup>RD</sup> pass</b> 50 - 59%	<b>Lower 2<sup>nd</sup> average</b> 60 – 69%	<b>Upper 2<sup>nd</sup> good</b> 70%	<b>1<sup>st</sup> very good –excellent</b> 75% +
<ul style="list-style-type: none"> <li>• Does not answer the question.</li> <li>• Main concepts absent</li> <li>• Off the point</li> </ul>	<ul style="list-style-type: none"> <li>• Struggles to answer the question</li> <li>• More than half of main concepts absent</li> <li>• Very little understanding of relevant concepts</li> <li>• Some accuracy in concepts and factual information</li> <li>• Very little accuracy in using facts and applying concepts</li> <li>• No reflection evident</li> <li>• No integration of different aspects of content and reflection</li> <li>• Hardly any use made of source material</li> <li>• Omitting most sources</li> </ul>	<ul style="list-style-type: none"> <li>• Question partially answered</li> <li>• Answer focuses on some concepts and neglects the others</li> <li>• Superficial understanding of relevant concepts</li> <li>• Some inaccuracy in concepts and factual information</li> <li>• Some inaccuracy in using facts and applying concepts</li> <li>• Instead of reflection, brief description of events or experiences (self and/or others' thoughts, feelings, behaviour) - lacks depth</li> <li>• Description confuses thoughts and feelings</li> <li>• Poor integration of different aspects of content and reflection</li> <li>• Little use made of source material</li> <li>• Selected some sources at expense of others</li> </ul>	<ul style="list-style-type: none"> <li>• Question fully answered</li> <li>• Answer includes all relevant concepts</li> <li>• Adequate understanding of relevant concepts</li> <li>• Accuracy in concepts and factual information</li> <li>• Mostly accurate in using facts and applying concepts</li> <li>• Reflective description of events and experiences (self and others' thoughts, feelings, behaviour) -some depth</li> <li>• Begins to link new and prior knowledge and/or experience</li> <li>• Begins to differentiate between feelings and thoughts</li> <li>• Adequate integration of different aspects of content and reflection</li> <li>• Adequate use made of all source material</li> </ul>	<ul style="list-style-type: none"> <li>• Question fully answered</li> <li>• Answer includes all relevant concepts</li> <li>• Good understanding of relevant concepts</li> <li>• Accuracy in concepts and factual information</li> <li>• Accuracy in using facts and applying concepts</li> <li>• Rich reflective description (self and others' thoughts, feelings, behaviour), shows relationship between issues and broader context</li> <li>• Links new and prior knowledge and/or experience</li> <li>• Differentiates between feelings and thoughts</li> <li>• Acknowledges different viewpoints</li> <li>• Good integration of different aspects of content and reflection</li> <li>• Appropriate use of source material</li> </ul>	<ul style="list-style-type: none"> <li>• Question fully answered</li> <li>• Answer includes all relevant concepts (and possibly other relevant concepts)</li> <li>• Excellent understanding of relevant concepts</li> <li>• Accuracy in concepts and factual information</li> <li>• Accuracy in using facts and applying concepts</li> <li>• In addition to rich description (self and others' thoughts, feelings, behaviour)etc, begins to show ability to reason and evaluate</li> <li>• Links new and prior knowledge and/or experience</li> <li>• Differentiates between feelings and thoughts, explains importance of identifying feeling in learning situation and explores impact of feeling on learning and behaviour</li> <li>• Discusses views from multiple perspectives</li> <li>• Excellent integration of different aspects of content and reflection</li> <li>• Excellent use made of all prescribed source material and possibly including further sources</li> </ul>

**Appendix 4**

**CONSENT FORM**

I hereby agree to participate in the research being conducted by Lorna Olckers of the School of Public Health at the University of Cape Town.

I understand that my name and identity will remain confidential, and that information gathered from me will be used as part of Lorna Olckers' Masters Thesis research.

I agree to be tape-recorded and/ or video-taped for the purposes of this research.

Signature.....

Name.....

Date.....

University of Cape Town

## **Appendix 5**

### **Interview Questionnaire**

1. Tell me about the process you go through when you are going to mark and when you are actually involved in marking the essay assignments.
2. What strengths do you think you bring to the assessment process?
3. Following on from the previous question, what challenges have you experienced in terms of your own competencies?
4. What importance, if any, do you attach to the fact that you are only allowed to mark essays from students from groups other than your own? What have been the implications of this particular marking process for you and your own students?

University of Cape Town

## **Appendix 6 Instructions for markers**

18 March 2005

Dear BP Markers

Thank you for agreeing to participate in my Masters Thesis study.

### **Context of the study:**

In 2002 and 2003 all BP and BHP “in-course performance based” essay assessments were graded by the students’ own facilitators, marked by a second marker and then moderated in a session in which any difficulties were discussed, and consensus finally reached. A number of difficulties with this process were clearly evident.

Marks assigned by the students’ own facilitators were frequently challenged in moderation as they were described by some markers as “subjective”.

Linked to the previous difficulty, was that some facilitators found the essay marking process very difficult. They struggled to shift roles from group facilitator to student assessor. This added significant stress to the job of some of the small group facilitators.

Some of the markers also struggled to give detailed written feedback to the students. The written feedback was therefore often not qualitatively consistent.

A significant amount of time was put into the training of facilitators as to how to assess written work. This was found to be extremely resource intensive in terms of it being time-consuming and expensive. With the natural turnover in staff in part-time positions, training had to be repeated and the new markers typically took longer than the more experienced facilitators to mark the essays.

In terms of the validity of scores, students questioned the assigned marks in terms of their fairness or reliability after comparing marks with their friends. I, as course convenor, sometimes found it difficult to defend the facilitators’ marks as they were clearly not always consistent between markers and with the marking criteria.

The 2002 and 2003 BP and BHP marking system and the resulting difficulties brought into question issues around the validity of the BP and BHP assessment process.

In 2004, a new system of marking was negotiated and implemented in response to the difficulties described above. An assessment team was put together consisting of facilitators with a particular interest in and commitment to assessing written work. The idea was that the validity concerns related to “subjectivity”, reliability of marks and feedback, fairness, facilitators who found the process stressful and challenging, and the expense and time-consuming nature of the previous system would be appropriately addressed through this new system.

With these new processes in place, the perception is that validity of the assessment tasks related to BP and BHP has improved. *But*, is this the case? Hence the study in which you are now participating.

**Thesis study plan:**

Friday 18 March

- Students hands in two copies of their first essay assignment
- The marking team is divided into pairs.
- Each individual in the marking pair receives papers belonging to his/her own student group and papers from their partner's group as well as extras from those students whose facilitator is not a marker.

Wednesday 30 March (by 12 noon)

- Markers send their individual marks to Lorna

Friday 1 April

- Moderation - Markers finalise marks - Alix enters these on the system.
- Lorna videos selected pairs.

Week of 4 to 8 April

- Essays are handed back to students.
- Lorna requests interviews with certain markers.

***Important instruction:***

***Please do not discuss or compare marks with anyone until after these have been submitted to Lorna. Discussion and appropriate changes to marks will happen in the moderation session.***

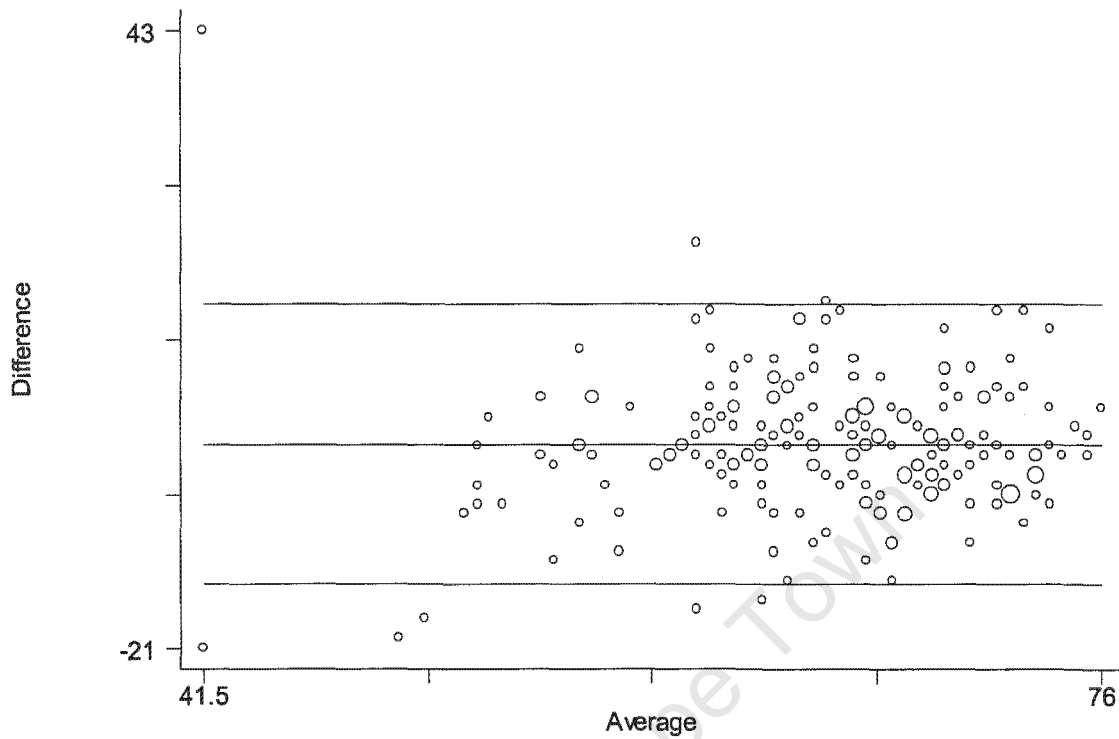
If you have queries, please feel free to call me.

Regards and thanks

Lorna Olckers

University of Cape Town

**Appendix 7**  
**Comparison of all markers 1 against all the markers 2**



	Marker 1:	Marker 2:
n	208	208
mean	65.5	65.3
Std. Dev	7.1	7

Bland-Altman comparison of marker1 and marker2  
 Limits of agreement (Reference Range for difference): -14.302 to 14.649  
 Mean difference: 0.173 (CI -0.816 to 1.162)  
 Range : 41.500 to 76.000

Paired t test

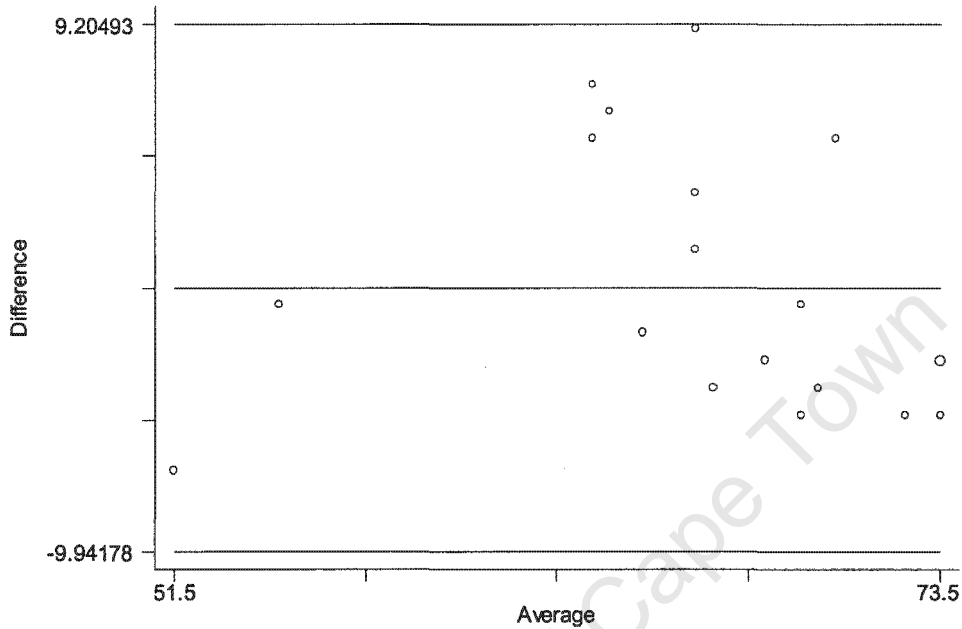
P = 0.7305

Results show no significant difference

## Appendix 8

Comparison of marks assigned by the student's own facilitator (marker 1) as compared with those assigned by another facilitator marker 2)

Group 1



	Marker 1: Anna	Marker 2: Tamsin
n	19	19
mean	66.6	67
Std. Dev	6.1	6.6

Bland-Altman comparison of marker1 and marker2

Limits of agreement (Reference Range for difference): -9.942 to 9.205

Mean difference: -0.368 (CI -2.676 to 1.939)

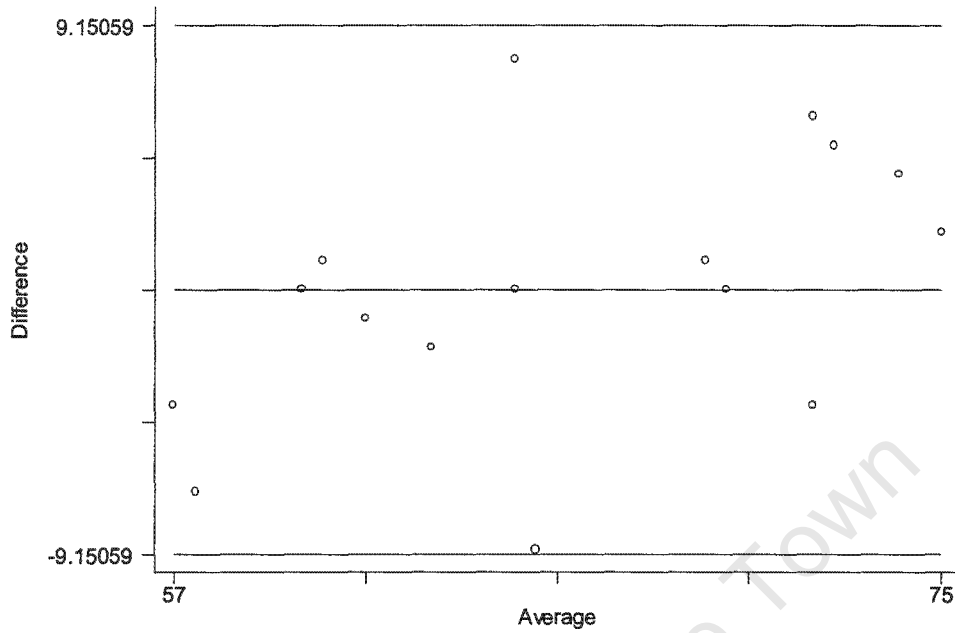
Range : 51.500 to 73.500

Paired t test

P = 0.7411

Results show no significant difference

Group 2



	Marker 1: Chloe	Marker 2: Jean
n	16	16
mean	66.3	66.3
Std. Dev	7.4	5.3

Bland-Altman comparison of marker1 and marker2

Limits of agreement (Reference Range for difference): -9.151 to 9.151

Mean difference: 0.000 (CI -2.438 to 2.438)

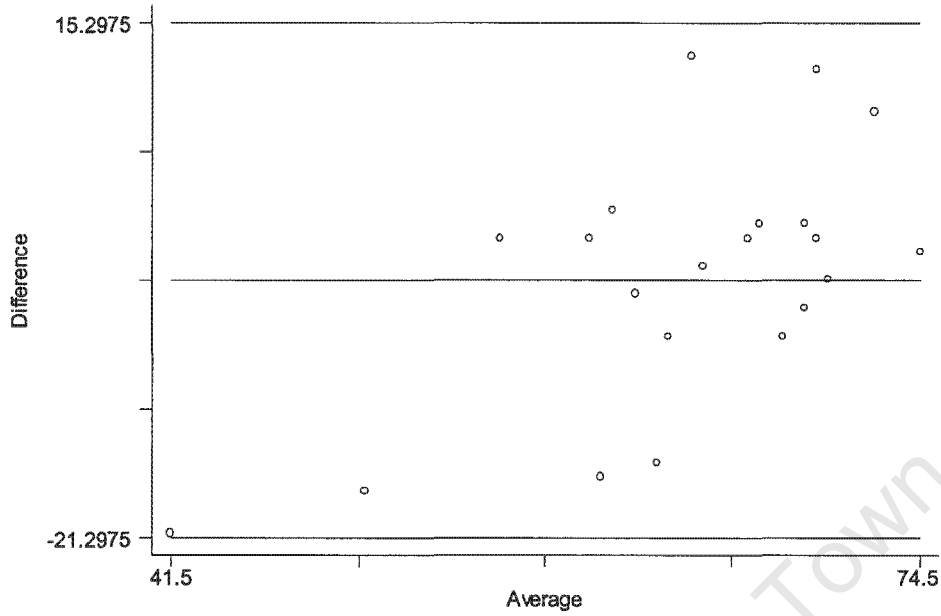
Range : 57.000 to 75.000

Paired t test

P = 1.0000

Results show no significant difference

Group 3



	Marker 1: Lea	Marker 2: Janet
n	21	21
mean	62.6	65.6
Std. Dev	11.2	6

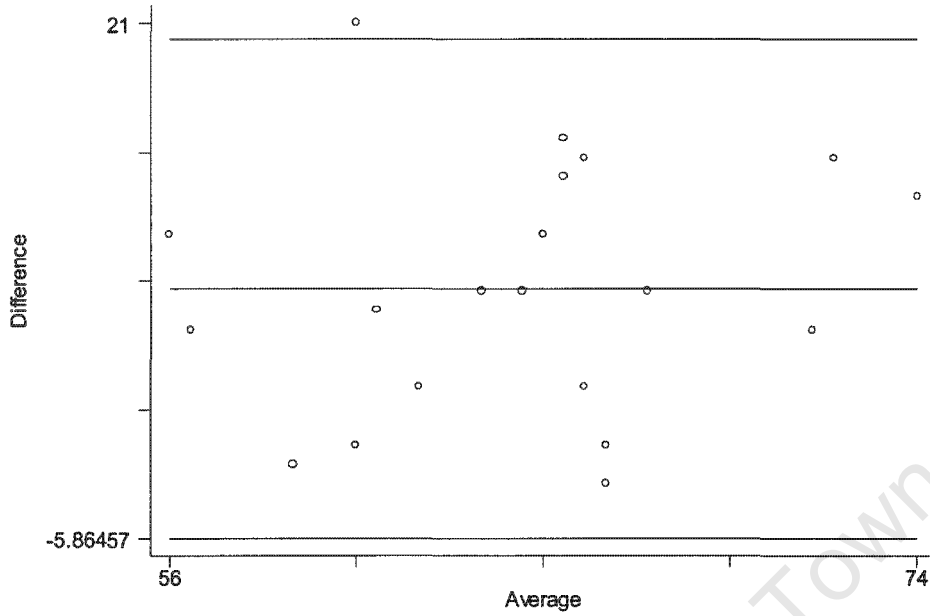
Bland-Altman comparison of marker1 and marker2  
 Limits of agreement (Reference Range for difference): -21.298 to 15.298  
 Mean difference: -3.000 (CI -7.164 to 1.164)  
 Range : 41.500 to 74.500

Paired t test

P = 0.1485

Results show no significant difference

Group 4



	Marker 1: Nicole	Marker 2: Tanya
n	20	20
mean	68	60.9
Std. Dev	6.2	5.3

Bland-Altman comparison of marker1 and marker2

Limits of agreement (Reference Range for difference): -5.865 to 20.165

Mean difference: 7.150 (CI 4.104 to 10.196)

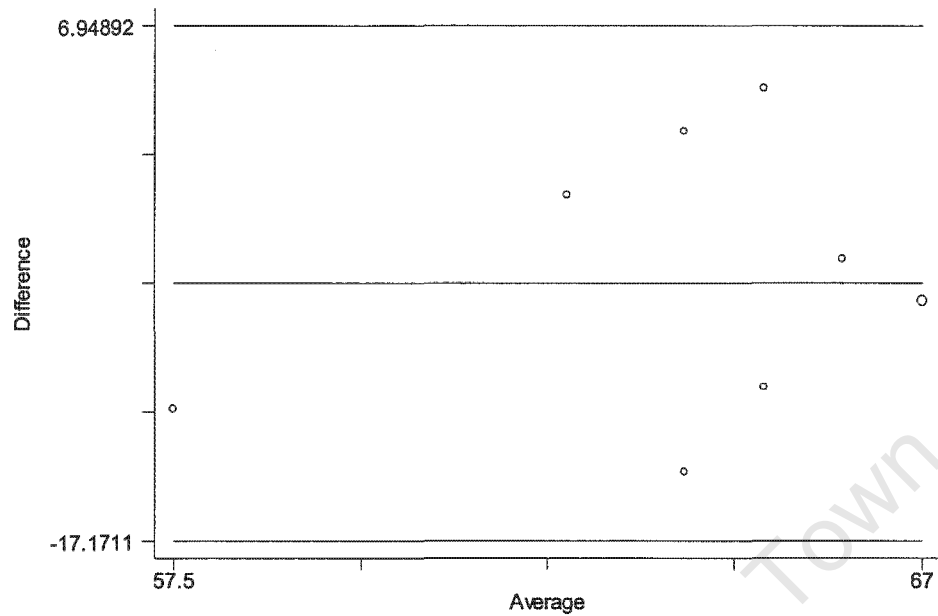
Range : 56.000 to 74.000

Paired t test

P = 0.0001

Results show a significant difference

Group 5



	Marker 1: Rene	Marker 2: Gail
n	9	9
mean	61.7	66.8
Std. Dev	4.7	3.7

Bland-Altman comparison of marker1 and marker2

Limits of agreement (Reference Range for difference): -17.171 to 6.949

Mean difference: -5.111 (CI -9.746 to -0.476)

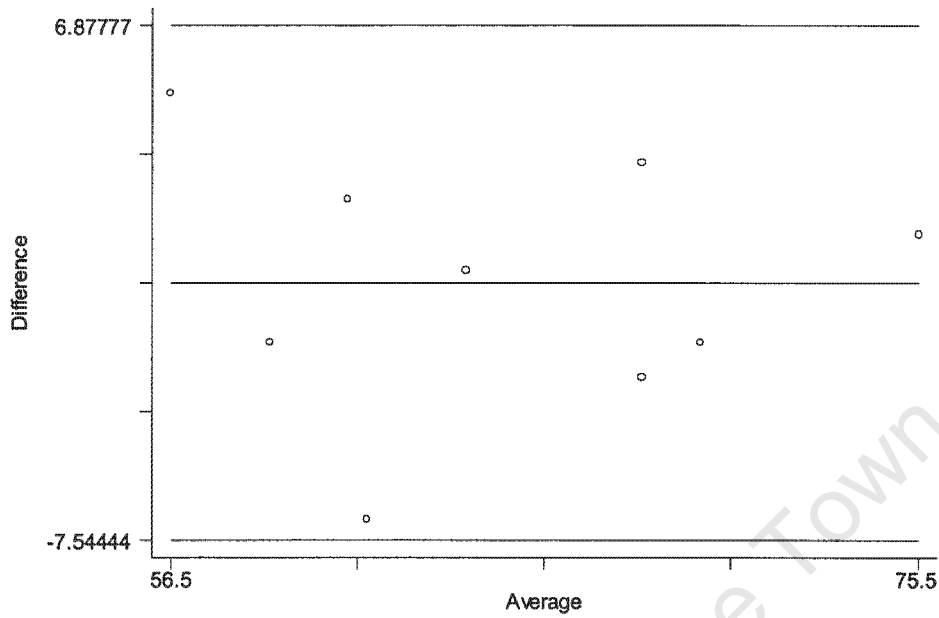
Range : 57.500 to 67.000

Paired t test

P = 0.0346

Results show a significant difference

Group 6



	Marker 1: Alfred	Marker 2: Charles
n	9	9
mean	64.8	65.1
Std. Dev	6.2	6.5

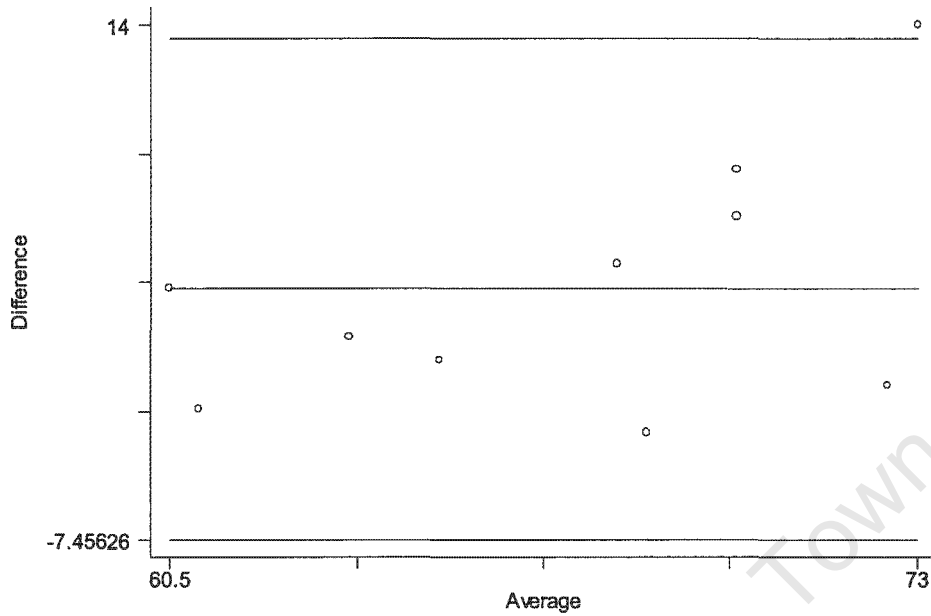
Bland-Altman comparison of marker1 and marker2  
 Limits of agreement (Reference Range for difference): -7.544 to 6.878  
 Mean difference: -0.333 (CI -3.105 to 2.438)  
 Range : 56.500 to 75.500

Paired t test

P = 0.7885

Results show no significant difference

Group 7



	Marker 1: Mary	Marker 2: Betty
n	10	10
mean	68.7	65.7
Std. Dev	6.2	3.9

Bland-Altman comparison of marker1 and marker2

Limits of agreement (Reference Range for difference): -7.456 to 13.456

Mean difference: 3.000 (CI -0.740 to 6.740)

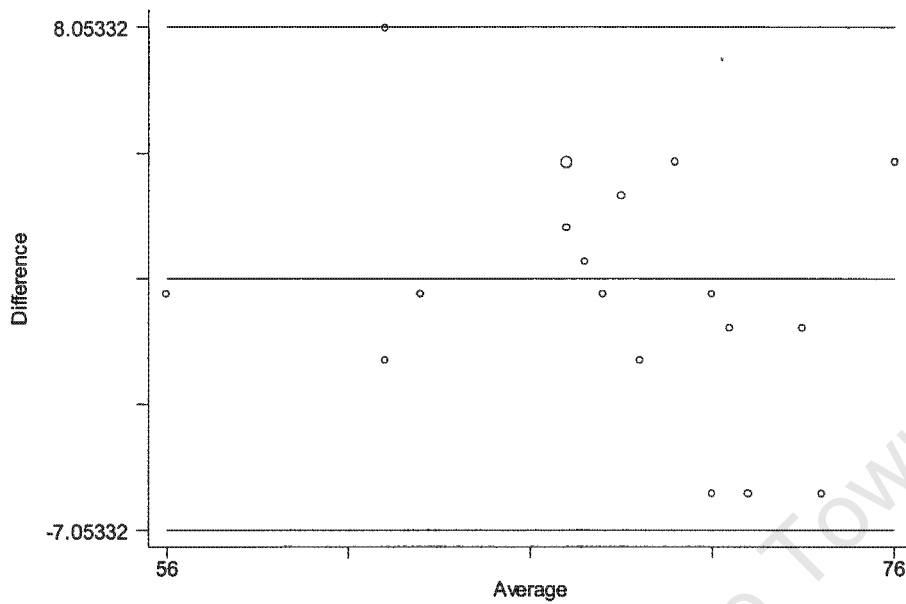
Range : 60.500 to 73.000

Paired t test

P = 0.1030

Results show no significant difference

Group 8



	Marker 1: Tamsin	Marker 2: Anna
n	20	20
mean	68.4	67.9
Std. Dev	4.6	5.6

Bland-Altman comparison of marker1 and marker2

Limits of agreement (Reference Range for difference): -7.053 to 8.053

Mean difference: 0.500 (CI -1.268 to 2.268)

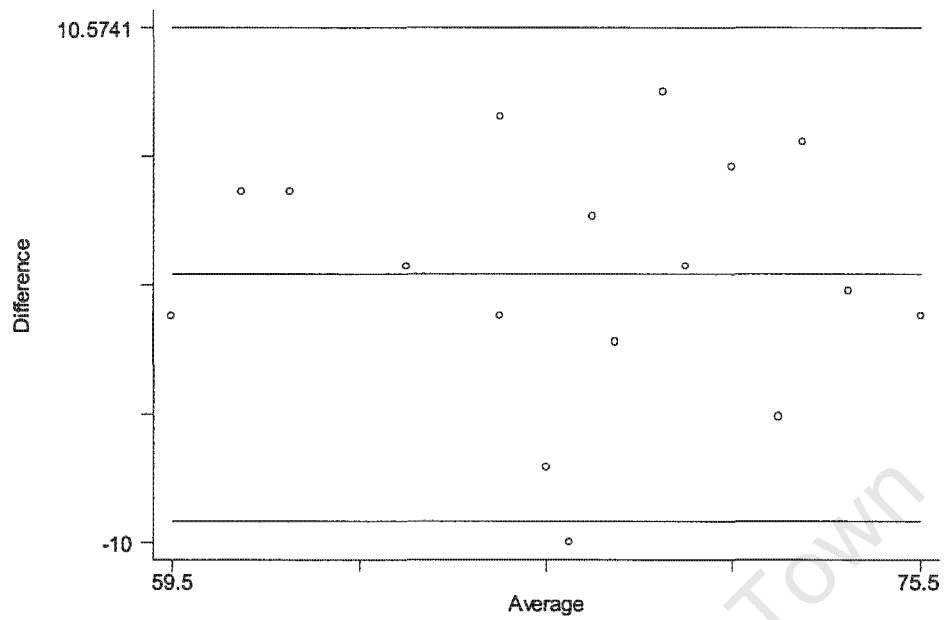
Range : 56.000 to 76.000

Paired t test

P = 0.5608

Results show no significant difference

Group 9



	Marker 1: Jean	Marker 2: Chloe
n	17	17
mean	68.6	67.9
Std. Dev	5	5.3

Bland-Altman comparison of marker1 and marker2

Limits of agreement (Reference Range for difference): -9.162 to 10.574

Mean difference: 0.706 (CI -1.831 to 3.243)

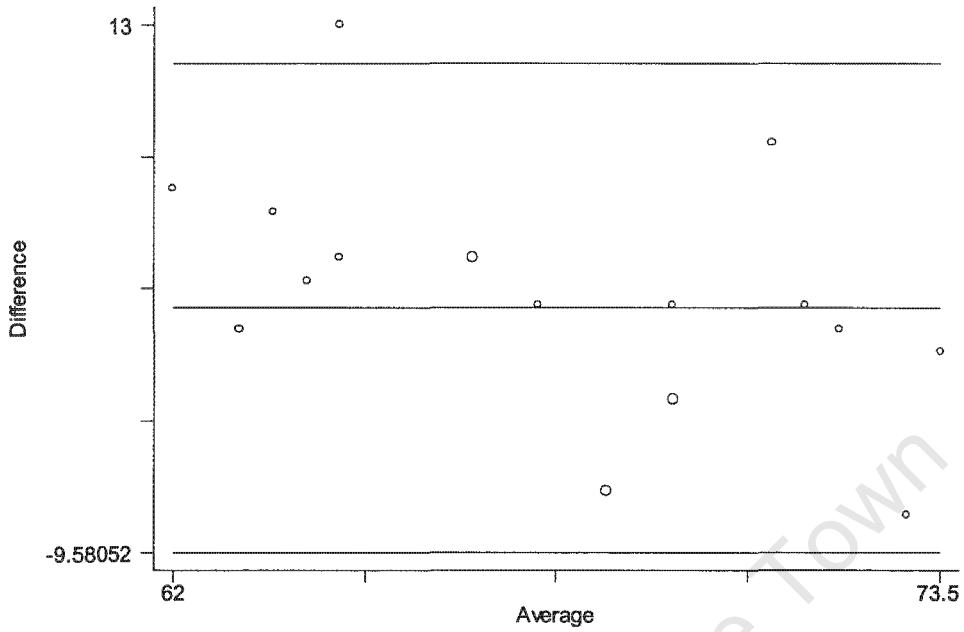
Range : 59.500 to 75.500

Paired t test

P = 0.5635

Results show no significant difference

Group 10



	Marker 1: Janet	Marker 2: Lea
n	19	19
mean	68.3	67.4
Std. Dev	3.2	5.3

Bland-Altman comparison of marker1 and marker2

Limits of agreement (Reference Range for difference): -9.581 to 11.370

Mean difference: 0.895 (CI -1.630 to 3.419)

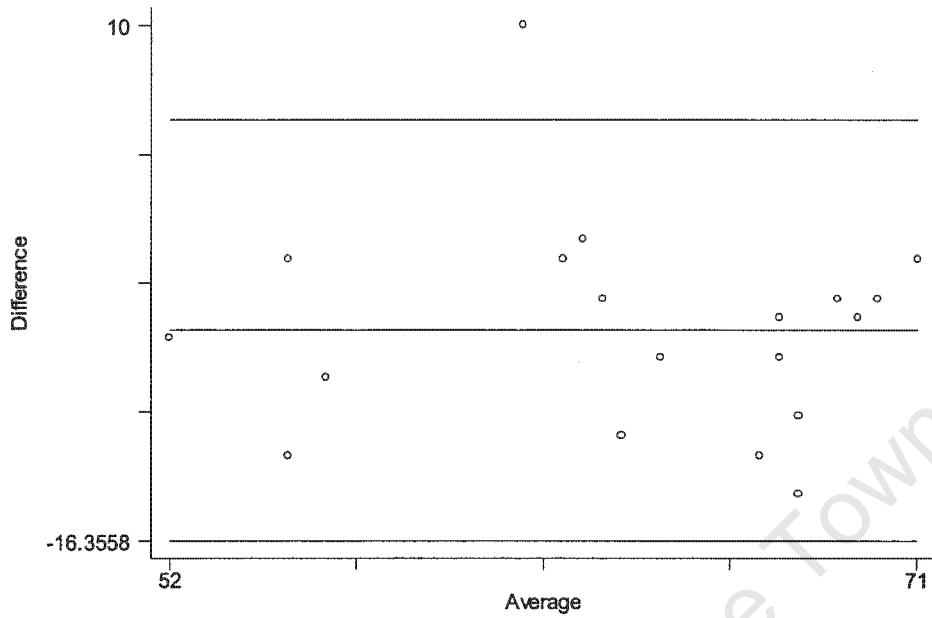
Range : 62.000 to 73.500

Paired t test

P = 0.4661

Results show no significant difference

Group 11



Bland-Altman comparison of marker1 and marker2  
 Limits of agreement (Reference Range for difference): -16.356 to 5.198  
 Mean difference: -5.579 (CI -8.176 to -2.982)  
 Range : 52.000 to 71.000

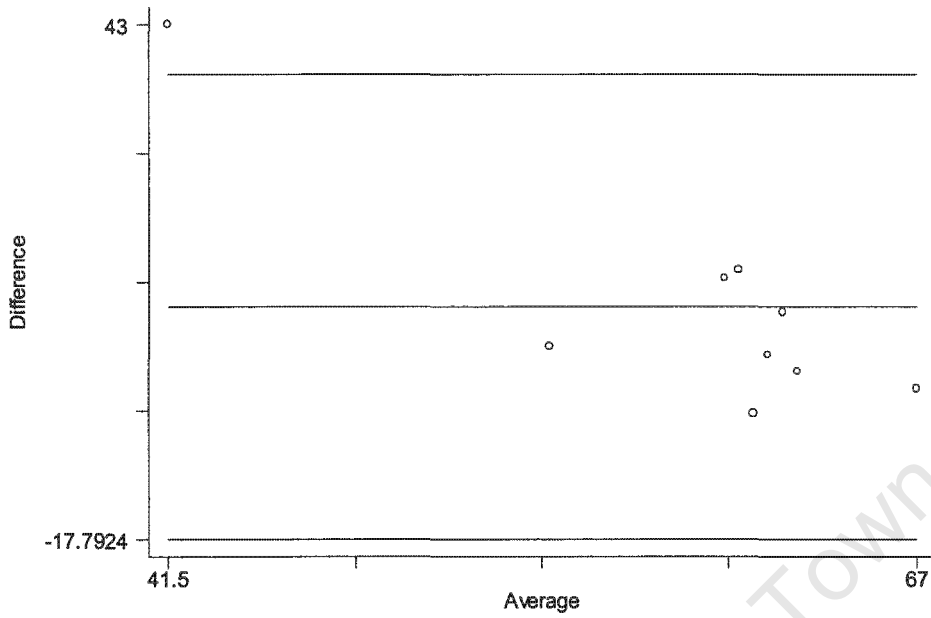
	Marker 1: Tanya	Marker 2: Nicole
n	19	19
mean	61	66.6
Std. Dev	6.2	6.5

Paired t test

P = 0.0003

Results show a significant difference

Group 12



	Marker 1: Gail	Marker 2: Rene
n	9	9
mean	64.1	54.4
Std. Dev	3.7	13.8

Bland-Altman comparison of marker1 and marker2

Limits of agreement (Reference Range for difference): -17.792 to 37.126

Mean difference: 9.667 (CI -0.887 to 20.220)

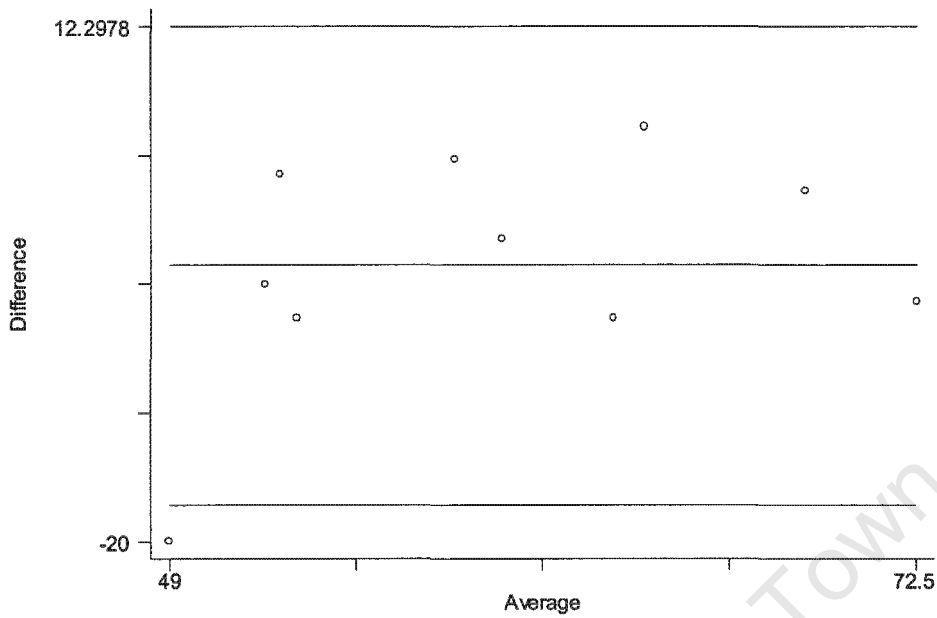
Range : 41.500 to 67.000

Paired t test

P = 0.0676

Results, although close, show no significant difference

Group 13



	Marker 1: Charles	Marker 2: Alfred
n	10	10
mean	57.9	60.6
Std. Dev	9.9	7.2

Bland-Altman comparison of marker1 and marker2

Limits of agreement (Reference Range for difference): -17.698 to 12.298

Mean difference: -2.700 (CI -8.064 to 2.664)

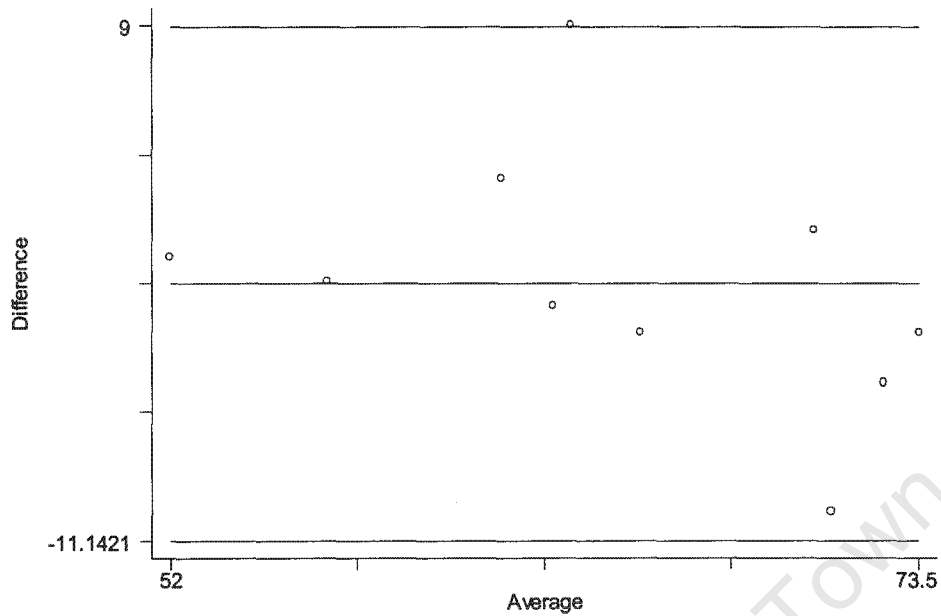
Range : 49.000 to 72.500

Paired t test

P = 0.2843

Results show no significant difference

Group 14



	Marker 1: Betty	Marker 2: Mary
n	10	10
mean	64.4	65.5
Std. Dev	6.5	8.4

Bland-Altman comparison of marker1 and marker2  
 Limits of agreement (Reference Range for difference): -11.142 to 8.942  
 Mean difference: -1.100 (CI -4.692 to 2.492)  
 Range : 52.000 to 73.500

Paired t test

P = 0.5059

Results show no significant difference

## REFERENCES

- Bernstein, R.J. 1983. *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis*. Philadelphia: University of Pennsylvania Press.
- Biggs, J. 1996a. Assessing learning quality: reconciling institutional, staff and educational demands. *Assessment and Evaluation in Higher Education*. 21(1):5–16.
- Biggs, J. 1996b. Enhancing teaching through constructive alignment. *Higher Education*. 32:347-364.
- Broad, B. 2003. *What we really value beyond rubrics in teaching and assessing writing*. Logan: Utah State University Press.
- Branthwaite, A., Trueman, M. & Berrisford, T. 1981. Unreliability of marking: further evidence and a possible explanation. *Educational Review*, 33(1) February: 41– 46.
- Burns, R. 1997. *Introduction to Research Methods*. 3<sup>rd</sup> ed. Australia: Addison Wesley Longman Pty Ltd.
- Crooks, T. & Kane, M. 1996. Threats to the valid use of assessments. *Assessment in Education: Principles, Policy and Practice*, 3(3): 265.
- Fraenkel, J.R. & Wallen, N.E. 2000. *How to design and evaluate research in education*. 4<sup>th</sup> ed. Boston: McGraw Hill.
- Gadamer, H.G. 2003. *Truth and Method*. 2<sup>nd</sup> ed. revised. New York: Continuum Publishing Company.
- Gilliant, S.E. & Hayward, N.F. 1996. A testing time: the role of subjective practices in making sense of student performance. *Assessment and Evaluation in Higher Education*. June 21(2): 161-172.

- Gibbs, G. 1999. Using assessment strategically to change the way students learn. In S. Brown Ed. *Assessment matters in Higher Education: Choosing and Using Diverse Approaches*. Buckingham: Society for Research into Higher Education and Open University Press.
- Gipps, C. 1999. Socio-cultural aspects of assessment. *Review of research in education/AERE*. 24: 355–392.
- Graaff, J., Reed, Y. & Shay, S. 2004. Validating academic assessment: a hermeneutical perspective. *Journal of Education*. 33: 51–68.
- Holmes, L. & Smith, L. 2003. Student evaluations of faculty grading methods. *Journal of Education for Business*. July/August 78 issue 6:318-323.
- Jervolino, D. 1996. Gadamer and Ricoeur on the hermeneutics of praxis. In Kearney, R. Ed. *Paul Ricoeur The Hermeneutics of Action*. London: Sage Publications.
- Kearney, R. Ed. 1996. *Paul Ricoeur The hermeneutics of action*. London: Sage Publications.
- Kirkwood, B.R. & Sterne, J.A.C. 2003. *Essential medical statistics*. 2<sup>nd</sup> ed. Massachusetts: Blackwell Publishing.
- Knight, P. 2002. The Achilles Heel of Quality: the assessment of student learning. *Quality in Higher Education*. 8(1):107–115.
- Kvale, S. 1996. *Interviews – an introduction to qualitative research interviewing*. California: Sage Publications.
- Light, G. & Cox, R. 2001. Assessing: Student Assessment. In *Learning and Teaching in Higher Education: The Reflective Professional*. London: Paul Chapman Publishing.

- Luckett, K. & Sutherland, L. 2000. Assessment practices that improve teaching and learning. In Makoni, S. Ed. *Improving Teaching and Learning in Higher Education*. Johannesburg: Witwatersrand University Press.
- McMillan, J.H. 1997. *Classroom Assessment: Principles and Practice for Effective Instruction*. Boston: Allyn and Bacon.
- Moss, P.A. 1994. Can there be validity without reliability? *American Educational Research Association*. 23 Part 2: 5–12.
- Moss, P.A. 1996. Enlarging the Dialogue in Educational Measurement: Voices from Interpretive Research Traditions. *Educational Researcher*. 25(1): 20-28.
- Nitko, A.J. 2001. *Educational Assessment of Students* .3<sup>rd</sup> Ed. New Jersey: Prentice Hall Inc.
- Pennycook, A. 1996. Incommensurable Discourses? *Applied Linguistics*. 15(2):115-138.
- Reagan, C.E. 1996. *Paul Ricoeur His life and his work*. London: University of Chicago Press Ltd.
- Reed, Y. and others. 2003. (Un)reliable assessment: A case study. *Perspectives in Education*, 21(1) March:15 – 28.
- Shay, S.B. 2003. The Assessment of Undergraduate Final Year Projects: a study of academic professional judgement. *Thesis for the Degree of Doctor of Philosophy*. Department of Education. University of Cape Town.
- Shay, S.B. 2004. The Assessment of Complex Performance: A Socially Situated Interpretive Act. *Harvard Educational Review*. 74(3).
- Shay, S.B. 2005. The assessment of complex tasks: a double reading. *Studies in Higher Education*. 30(6): 663-679.

- Shepard, L. 2000. The role of assessment in a learning culture. *Educational Researcher* 29(7): 4–14.
- Thompson, J.B.Ed. 1981. *Paul Ricoeur - Hermeneutics and the human sciences*. New York: Cambridge University Press.
- Torrance, H. 2000. Postmodernism and Educational Assessment. In Filer, A. *Assessment: Social Practice and Social Product*. London: Routledge and Falmer.
- Wenger, E. 1998. *Communities of Practice learning, meaning, and identity* United Kingdom: Cambridge University Press.
- Wyatt-Smith, C. & Castleton, G. 2005. Examining how teachers judge student writing: an Australian case study. *Journal of Curriculum Studies*. 37(2) 131-154.

University of Cape Town