Assessment and the quality of educational programmes: What constitutes evidence?

SUELLEN SHAY AND JEFF JAWITZ

SUELLEN SHAY *is a senior lecturer in the Centre for Higher Education Development (CHED) at the University of Cape Town. She heads the Academic Staff Development Group of CHED and convenes the Higher Education Studies post-graduate programme. Her current research interest is educational assessment in higher education.*

JEFF JAWITZ *is a senior lecturer in the Academic Staff Development Group and the Academic Development Programme at the University of Cape Town. He convenes a Masters course on Learning and Teaching in Higher Education and a programme to help new academic staff in developing as researchers and educators. His current research interest is in academic staff development in higher education.*

Abstract

In a climate of growing accountability for Higher Education, there is an increased demand on assessment to play an evaluative role. National, professional and institutional quality assurance systems expect that the assessment of student performance can be used to evaluate the quality of teachers, learners, programmes and even institutions for the purposes of programme review, programme accreditation and institutional audits. While affirming the role of assessment as a catalyst for the improvement of teaching and learning, this paper problematises the use of assessment for evaluative purposes. Drawing on an Engineering Council of South Africa (ECSA) accreditation exercise, we illustrate the limitations of particular forms of assessment data as evidence of quality. The paper argues for assessment as a socially situated interpretive act and for validation as the on-going process of strengthening the alignment between our assessment tasks, procedures and outcomes, and the educational, political and social purposes which our assessment systems claim to serve. The central argument of the paper is that key source of assessment evidence for quality assurance purposes is to be found in the strength and effectiveness of programme validation systems.

Introduction

In addition to the multiple purposes which educational assessment has traditionally served, for example diagnostic, summative and formative, there is increasing reference in the literature to assessment's evaluative purpose (Lambert & Lines, 2000; Luckett & Sutherland, 2000; Weedon, Winter & Broadfoot, 2002). This suggests that through the assessment of student performance the quality of teachers, learners, programmes and institutions can be evaluated for external purposes, for example for programme review, programme accreditation and institutional audit. If, however, assessment is to serve as a 'window' on the quality of teaching and learning, what counts as *evidence* of quality? This was one of the key questions posed at the American Association of Higher Education's 2004 Assessment Conference. Assessment scholars and practitioners throughout North America gathered to consider how assessment can strengthen higher education's capacity to respond to the growing demands for accountability with respect to the quality of teaching and learning. Despite the lively debate, there seemed to be no consensus on what kinds of evidence can be generated from assessment systems for quality assurance purposes.

As educational developers in higher education both authors have been involved in the development and research of assessment systems across a variety of disciplines. We also have experience with quality assurance exercises at a programme and institutional level. One of the authors has been involved in helping colleagues prepare for the Engineering Council of South Africa (ECSA) accreditation process. In this paper we draw on these experiences to explore the opportunities and limitations of assessment's evaluative role. Through the ECSA case study we illustrate the limitations of assessment data (e.g. tasks, performances, outcomes) as evidence of quality, and critique assumptions concerning the role that assessment can play in measuring the quality of graduates for programme accreditation purposes. Our critique calls for a re-conceptualisation of assessment and validity, and introduces the notion of a programme validation system. We argue that it is methodologically problematic to draw on assessment data as evidence for evaluative purposes. Instead evidence for the quality of teaching and learning is to be found in the effectiveness of validation systems. We close by giving an example of a validation mechanism for final-year projects in an engineering faculty.

Assessment: Catalyst for reform and tool for accountability

The assessment literature of the past two decades has been dominated by two themes. The first theme is assessment as a catalyst for educational reform (or improvement); the second more recent theme is assessment as a technology for accountability (or quality assurance) (Linn, 1993). With reference to the educational reform agenda, assessment is considered to be one of the most powerful influences on what and how teachers teach, and what and how students learn. Thus it follows that improvement in classroom assessment will make a strong contribution to the improvement of learning (Black & Dylan, 1998, Ramsden, 2003). On the basis of this key principle assessment scholars have argued that "the quickest way to change student learning is to change the assessment system" (Elton & Laurillard cited in Biggs, 1996, 5); that given assessment's capacity to shape the curriculum, "assessment reform is the most urgent priority confronting undergraduate education" (Brown & Knight, 1994, 11-12).

As the demand upon higher education for greater accountability has increased, assessment has been perceived to be a powerful way of measuring the quality of teaching and learning, as well as the quality of teachers and learners. Thus assessment is perceived not only as an instrument of reform, but as one of the primary technologies for measuring and reporting on that reform (Broadfoot, 1996, 1999). This has raised a great deal of optimism among both educators and policy-makers that assessment will be able to assist higher education institutions to demonstrate "exactly" what the state "gets" for its "enormous investment" (Brown & Knight, 1994, 12).

Consistent with these global trends, national quality assurance systems in South Africa are making explicit demands on higher education assessment systems in the interests of improving and assuring the quality of teaching and learning. The coupling of assessment with the quality assurance agenda is well illustrated in South Africa's National Qualifications Framework (NQF) where learning outcomes and criterion-based assessment serve as one of the technologies for rectifying the inequalities of the past by facilitating access, mobility and progression through a qualification system which was previously fragmented along racial, political and geographic lines. The explicit focus on assessment in the Higher Education Quality Council's (HEQC) criteria for institutional audit and programme accreditation further points to the role allocated to assessment in the national quality assurance system for higher education. Assessment practices are considered to be a key indicator of the 'health' of teaching and learning and thus a legitimate focus of quality assurance activities (HEQC, 2003).

On the basis of our experience we strongly support the notion of assessment as a catalyst for the improvement of teaching and learning. We agree with those cited earlier that assessment is potentially a powerful point of leverage on the quality of teaching and learning, although the challenges of shifting assessment practices should not be underestimated (Biggs, 1996). We are therefore supportive of the attention given by the HEQC and the professional associations to assessment in higher education. While supporting the notion that assessment can *enhance* quality, we wish to problematise the notion that assessment can *measure* quality. We return to elaborate on this point below.

Using assessment for measuring quality

In 2001 ECSA conducted its 5-yearly accreditation of engineering undergraduate programmes at the University of Cape Town. It was the first accreditation visit using ECSA's newly adopted outcomes-based approach and represented a shift in focus from content and structure to whether engineering programmes ensured that their graduates satisfy a set of ten generic exit-level outcomes. In the new approach the ECSA accreditation team was required, amongst other things, to interrogate the engineering programme in relation to four questions evaluating the following areas: learning outcomes, content, the teaching and learning process, and sustainability of the programme respectively. For the purposes of this paper we focus only on the first question:

> Question 1: Does the assessment within the programme verify that every student satisfies the learning outcomes (as specified)? (ECSA 2000)

In the final report the accreditation team was required to answer Question 1 for each specified outcome with either a "yes" or a "no" with space available for the team to motivate their decision with reference to the evidence they had found. An example of such an outcome is:

> Outcome 2.1: The BSc (Eng) graduate is competent to identify, assess, formulate and solve convergent and divergent engineering problems creatively and innovatively (ECSA 2000).

The assumption underlying Question 1 is that assessment can provide evidence for the achievement of learning outcomes and that this evidence can contribute to a judgment on the quality of a programme and the quality of graduates. Our interest was in the kinds of evidence the programme team put forward to verify that every graduate satisfies the learning outcomes and the kinds of evidence the accreditation team drew on in order to reach a judgment on this question.

The programme teams are required to submit documentation six weeks prior to the visit. The ECSA guidelines stipulate that "each outcome specified … must be explicitly addressed in

terms of the *means of assessment* and the *criteria* for satisfaction of each outcome at exit level. The choice of evidence and format of presentation is left to the academic entity" (ECSA 2002). During the visit the ECSA accreditation team is expected to examine material available on site and elicit further information to be able to answer Question 1. Guidelines indicating what materials should be available during the visit refer to two types of assessment data:

i) Test papers, assignments and project statements, examination question papers with specimen solutions for the last three years in each course; and
ii) a sample of marked examination scripts for the most recent examination in each course, and final-year project and design reports "representative of … good, average and just passing/failing students".

In trying to understand what evidence the accreditation teams valued during the UCT visit in 2001, we examined the motivational commentary inserted in the final reports by the ECSA teams in response to Question 1 with reference to Outcome 2.1. In all eight programmes the accreditation teams had reached affirmative judgments. The written commentary provided some indication of the motivation for these decisions. Comments included the following: "Study of *exam questions* gave evidence of this as did *discussions with students and staff*" and "the outcome is supported by the following evidence: Sampling of *exam papers* … examining *design projects and problems* set during the various course years". While acknowledging that these comments are extremely limited and telegraphic, they would appear to indicate that the accreditation teams used physical samples of examination scripts, papers and projects, and discussions with staff and students to infer that the assessment in the programme verifies that every student achieves the required exit level outcome.

The ECSA accreditation process illustrates the use of assessment as one 'window' on the quality of graduates for purposes of programme accreditation. It also highlights the kinds of assessment data which programme teams put forward as evidence for whether "every student satisfies the learning outcomes". We now turn to explore the assumptions about what assessment is and the expectations of what it can deliver which underlie this use of assessment data.

Shifting assessment discourses

Definitions of assessment commonly refer to it as a process of collecting, analysing and interpreting information about learners' responses for a variety of different purposes (Lambert & Lines, 2000; Nitko, 2001; Gronlund, 1985). There have however been important shifts in assessment discourse (Gipps, 1999). In contrast to traditional understandings of assessment as an "instrument for measuring behavior" (Gronlund, 1985), the notion of assessment as an interpretive process is one which is gaining prominence in the literature (Gipps, 1999; Knight, 2002a and 2002b; Shay 2004). This shift in emphasis signals a departure from objectivist notions of assessment as an accurate, reliable scientific process of applying instruments to measure learner performance in order to quantify achievement. In contrast, there is a growing recognition that assessment, particularly of complex tasks, is a judgment which is socially situated and thus contingent on a variety of factors which constitute the assessment event – the learning context, the nature of the assessment task, the purposes of the assessment, and the relationship between the assessor and assessed (Shay, 2003). Thus assessment judgments are the outcome of complex interpretive processes.

The notion of assessment as a socially situated interpretive act, together with the complexity of learning processes valued in higher education, suggests that we need to be cautious about claims of assessment's ability to measure student learning for quality assurance purposes. Various assessment scholars have critiqued the use of assessment data for judgments on the quality of teachers, graduates, programmes or institutions (Knight, 2002a and 2002b;

Broadfoot, 1999; Davis, 1998; González & Burwood, 2003 in the UK; Berlack *et al.*, 1992; Pearson *et al.*, 2001, in the U.S). For example Knight (2002a and 2002b) problematises the use of grades and degree classifications in the UK as evidence for "high stakes purposes" given that these assessments are partial and potentially unreliable measures of achievement. He argues that assessment cannot serve these "managerialist agendas", and thus "radical thinking is needed about what summative assessment is for, who it is for, what it can do, what it cannot do cheaply and what it ought not to be asked to do at all" (Knight, 2002b). Fortunately, South Africa's higher education quality assurance system (as implemented by the HEQC) has not adopted the UK practice of using quantitative assessment data as performance indicators of quality, although the ECSA case above raises questions on the current use of assessment data such as the final-year project scripts as evidence of quality.

This shift in the notion of assessment has been accompanied by an associated shift in the concept of validity. Where assessment is understood as a measurement of sample behaviour, it follows that validity is conceptualised largely as a technical, quantitative operation for ensuring that "the test measures what it is meant to measure" (Hill in Killen 2003, 2). But with the shift to assessment as interpretation, validity scholars argue that validity is best understood as the extent to which assessment *interpretations* are sound for the purposes for which they are being used (Messick, 1989; Nitko, 2001). This understanding suggests that validity is not a fixed property of a particular assessment instrument or set of results, but always contingent upon the *interpretations* and the *purposes* for which those interpretations are used. Given this emphasis of validity as an evolving rather than fixed property, some validity scholars suggest that the preferable term is "validation" as it signals validity as an on-going *process* of strengthening the alignment between our assessment interpretations and the educational, political and social purposes which our institutions claim to serve (Cronbach, 1989; Messick, 1989; Mishler, 1990). This on-going validation requires mechanisms at institutional, faculty, programme and course level which ensure the soundness of our assessment interpretations for their uses. For example a key validation mechanism in the South African higher education system is external examination. Before we return to the central issue of what constitutes evidence of quality, a crucial prior question needs to be addressed: by what criteria do we judge the soundness of our interpretations for their uses?

Recent validity scholarship has given a great deal of attention to the criteria by which the soundness of assessment interpretations is to be judged (see Moss 1992 for an overview). This work on the criteria for validity is most often applied in the field of testing, in particular large-scale standardised testing for high stakes purposes. There has been less attention to the design and application of validity criteria for course and programme level assessment of student learning. Table 1 provides a list of criteria against which a course or programme can systematically interrogate its assessment system, and ultimately the soundness of the interpretations that emerge from the system. As noted above, current practice in higher education relies heavily on the external examination system for the validation of its interpretations. Historically external examination was established to ensure consistency across institutions (Warren Piper, 1996). While consistency (or reliability) is an important criterion this is only one kind of evidence of soundness. We would argue that alignment is the most important criterion, that is, the extent to which there is correspondence between what we value (the knowledge, skills and attitudes) and what we assess. The crucial issue therefore is ensuring that there are mechanisms which enable us to interrogate course and programme alignment (as well as the other criteria listed in the table) in an on-going way. These mechanisms at all levels of the institution (course, programme, department, faculty) are what constitute what we refer to as a validation system.

The argument in this article is that the key source of assessment evidence for quality assurance purposes is to be found in the strength and effectiveness of the validation system. Thus assessment can provide quality assurers with three kinds of evidence. The first is evidence

Table 1[1]: Validity criteria for course and programme level assessment

Alignment	There is a correspondence between course objectives/aims, the teaching activities (teaching methods and materials) and the assessment practices (methods, criteria, feedback).
Reliability	There is consistency in marking student performance: consistency within an individual marker, consistency between multiple markers in the same course, consistency within a department and Faculty.
Security	The necessary procedures are in place to ensure that opportunities for cheating are limited. Mechanisms are in place to ensure that where cheating has occurred it can be detected.
Transparency	Information about the assessment is made available to students. Students will be informed about: why they are being assessed, when they will be assessed, what methods will be used to assess, what criteria will be used to assess, how the final mark is derived, and appeal mechanisms.
Fairness	In the design and administration of the assessment there is sensitivity to issues of language and cultural diversity. Assessment administration ensures that students with particular disabilities are adequately catered for. Precaution has been taken to minimise conscious and unconscious discrimination for or against students on the basis of race, gender or any other form of prejudice.
Legitimacy	Both those who are being assessed and those who use the assessment data perceive the assessment to be an appropriate, fair and worthwhile exercise.
Consequences	There are mechanisms in place to monitor unintended consequences which may result from assessment interpretations and their uses.
Feasibility	The costs and practical considerations for administering the assessment are reasonable.
Administration	There are adequate human and material resources to ensure that assessment data is efficiently and accurately processed.

that there is a validation system in place with key mechanisms at all levels by which criteria (such as those described in Table 1) are reviewed in an on-going manner. The second is evidence of the effectiveness of these mechanisms in exposing both excellent and weak educational practice. The third is evidence that strong educational practice is being rewarded, and that weak educational practice is being addressed in constructive ways intended to lead to improvement. We would argue that the last evidence is ultimately the most indicative of quality teaching and learning.

We now turn to illustrate the development of a validation mechanism for engineering programmes at UCT and the way in which it exposed aspects of assessment which need improvement.

Developing a validation mechanism

Final-year projects form the culminating assessment event of undergraduate engineering programmes. They serve a formative role as a final integrative assessment opportunity, as well as a summative role in that their successful completion is a requirement for graduation. An additional purpose which is gaining momentum, as noted above, is their use for evaluative purposes, that is, as a measure of the exit-level quality of graduates, and hence the quality of the programme.

* These criteria have been extracted from the University of Cape Town's assessment policy which the authors of this paper were responsible for drafting.

Our own research on the assessment of final-year engineering projects revealed a variety of approaches and attitudes to assessment amongst academic staff with implications for the validity of the assessment system (Jawitz *et al*., 2002). As a result we developed a Management and Assessment Planning Protocol (MAPP) to help programmes develop a procedure for validating their assessment practice. Our work was informed by the approach that improvement in teaching practice is facilitated by helping staff develop the ability to reflect on their practice and to take action as a result of such reflection (Butler, 1996; Schön, 1987).

MAPP is designed to generate conversations amongst academic staff about their assessment practice by taking them through a sequence of questions covering the stages involved in the management and assessment of final-year projects (see Appendix 1 for an extract from MAPP). It facilitates the interrogation of existing assessment practices against a range of validity criteria and provides an opportunity for academic staff to explore collectively alternatives to their existing practice.

MAPP was developed and tested in two departments (A and B) with very different characteristics in terms of staff and student numbers, established procedures, and levels of collective engagement on key aspects of the final-year project. The use of assessment data for evaluative purposes assumes and indeed relies on a level of consensus and standardisation of purpose, practice and procedure. The discussion generated by MAPP revealed this assumption of standard practice as problematic and highlighted substantial differences in assessment practice and beliefs amongst academics in the same department.

For example, in department A staff differed substantially on specific issues such as the value and use of the official departmental set of marking criteria. While a new member of staff found the marking criteria helpful, a senior academic argued that it was a problem to have a standardised departmental "score sheet". How these criteria were interpreted and applied varied widely across supervisors. Academics in department B were found to be divided in their views on the overall purpose of the final year project. For some it served primarily as preparation for the workplace while others saw it as a culminating integrated assessment of the undergraduate degree. A third view regarded it as preparation for postgraduate studies. MAPP also revealed disagreements on a range of specific issues, such as what constitutes a valid project and how much assistance students should be given.

The existence of such differences in practice and belief within and across departments reinforces our concern about how the assessment of final-year project is used for evaluative purposes by outside stakeholders such as an ECSA accreditation team. However by making these inconsistencies visible, a validation procedure such as MAPP lays the groundwork for improving and strengthening the assessment practice.

In both departments MAPP served to facilitate the identification and prioritisation of aspects of assessment practice that needed improvement. In response to questions F3 to F5 (See Appendix 1) academics in department A recognised the limitations of their current marking memorandum and prioritised the development of clearer assessment criteria. In response to questions E1 and E4, they recommended that the guidelines for and preparation of supervisors of the final-year project be improved. In response to question F2 academics in department B agreed that the existing practice of only considering the written project report for the final mark should be changed to include the oral presentation.

However a validation system cannot stop at identifying what improvements are necessary – it must include a regular review of the implementation of these suggested improvements. We would argue that to ensure quality, the effectiveness of a validation mechanism such as MAPP needs to be monitored over several cycles to ensure that continuous improvement is built into programme management. It is only at this level of interrogation of the assessment practice that one can begin to make judgments on the quality of a programme.

MAPP serves to illustrate another theme which is noted in the validity scholarship but not sufficiently highlighted in practice. If we accept assessment as an interpretive process, then it follows that the validation of our assessment-based interpretations is a community process. It is only through the "community of inquirers" (markers, external examiners, professional bodies) who are "able, willing and committed to engage in argumentation" (Bernstein, 1976, 111) that we test the soundness of our interpretations for their uses. We believe that validation mechanisms such as MAPP contribute to the development of such communities and that it is the actions of these communities, committed to critical reflection and on-going improvement, that are ultimately the evidence of a quality teaching and learning.

Conclusion

In the context of growing demand for accountability, Weedon *et al.* (2002) pose the question, "is the assessment regime asking the right questions?" We would argue that the Question 1 ECSA currently uses is possibly not the right question. Underlying such a quest to "verify" are traces of objectivist notions of assessment as an accurate, reliable scientific process of measuring learner performance in order to quantify achievement. If we accept the assessment of complex tasks as a socially situated interpretive act, then this kind of question is not answerable. The notion that assessment can measure the achievement of complex learning outcomes for evaluative purposes is a "trick which assessment simply cannot bring off" (Davis, 1998, 7).

Neither does the question serve to generate the kind of activity that leads to the development of improved practice. Gosling and D'Andrea (2001) note that academics in the UK spend a great deal of time on activities such as collecting documentation that does not necessarily enhance the quality of programmes. They contrast quality *assurance* with quality *development* where the focus of the latter is on practice not documentation. We offer MAPP as one example of a quality development process through which academic staff are able to reflect on their assessment practices in ways which have the potential to lead to improved practice. However, as noted above, in the face of competing priorities within higher education, systemic changes in assessment practices are not easily implemented. These changes require institutional commitment to improve the quality of teaching and learning. We therefore suggest that a better question would be:

"What validation system does the programme have in place, and how does it contribute to improving teaching and learning?"

Such a question we feel would help focus quality assurance activities not only on the existence of validation systems but also on the commitment of the institution to see through and resource the recommendations for improved practices that arise out of such systems.

References

Berlack H, Newmann F, Adams E, Archbald T & Romberg T 1992. *Toward a new science of educational testing and assessment*. New York: State University of New York Press.

Biggs J 1996. Assessing learning quality: reconciling institutional, staff and educational demands. *Assessment and Evaluation in Higher Education*, **21**(1), 5-15.

Black P & Dylan W 1998. Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practices* **5**(1), x.

Broadfoot P 1996. *Education, assessment and society: A sociological analysis*. Buckingham: Open University Press.

Broadfoot P 1999. Empowerment or performativity? English assessment policy in the late twentieth century. Paper presented as part of the Assessment Reform Group Symposium on Assessment Policy presented at the British Educational Research Association Annual Conference, University of Sussex at Brighton.

Broadfoot P 2000. Preface. In A Filer (ed.), *Assessment: Social practice and social product.* London: Routledge and Falmer, ix-xii.

Brown S & Knight P 1994. *Assessing learners in higher education*. London: Kogan Page.

Butler J 1996. Professional development: Practice as text, reflection as process and self as locus. Australian Journal of Education, **40**(3), 265-283.

Cronbach L 1989. Construct validation after thirty years. In R Lynn (ed.), *Intelligence: Measurement, theory, and public policy: Proceedings of a symposium in honor of Lloyd G. Humphreys*. Urbana: University of Illinois Press, 146-171.

Davis A 1998. Special issue: The limits of educational assessment. *The Journal of the Philosophy of Education*, **32**(1), 1-150.

ECSA 2000. Standards for Accredited University Engineering Bachelors Degrees, PE-61 Rev 1, Engineering Council of South Africa.

ECSA. 2002. Documentation requirements for Accreditation Visits to Universities. PE-73 Rev 1. Engineering Council of South Africa.

Gipps C 1999. Socio-cultural aspects of assessment. *Review of research in education/ AERA*, **24**, 355-92.

Gosling D & D'Andrea V 2001. Quality development: A new concept in higher education. *Quality in Higher Education* **7**(1), 7-17.

Gonzalez AS & Burwood S 2003. Tacit knowledge and public accounts. *The Journal of the Philosophy of Education Society of Great Britain* **37**(3), 377-391.

Gronlund N 1985. *Measurement and evaluation in teaching*. New York: Macmillan Publishing Company.

Higher Education Quality Committee. 2003. Improving teaching and learning resource. Draft for 2003 workshops, Council on Higher Education.

Jawitz J, Moore R & Shay S 2002. Management and assessment of final year projects in engineering. *The International Journal of Engineering Education*, **18**(4), 472-478.

Killen R 2003. Validity in outcomes-based assessment. *Perspectives in Education*, **21**(1), 1-14.

Knight PT 2002a. The Achilles' heel of quality: The assessment of student learning. *Quality in Higher Education*, **8**(1), 107-15.

Knight PT 2002b. Summative Assessment in Higher Education: Practices in disarray. *Studies in Higher Education*, **27**(3), 275-86.

Lambert D & Lines D 2000. *Understanding assessment: Purposes, perceptions, practices.* London: Routledge Farmer.

Linn RL 1993. Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis*, **15**(1), 1-16.

Luckett K & Sutherland L 2000. Assessment practices that improve teaching and learning. In S Makoni (ed.), *Improving teaching and learning in higher education*. Johannesburg: Witwatersrand University Press, 98-130.

Messick S 1989. Meaning and values in text validation: The science and ethics of assessment. *Educational Researcher*, **18**(2), 5-11.

Mishler E 1990. Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Education Review*, **60**(4), 415-42.

Moss PA 1992. Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, **62**(2), 229-58.

Nitko A 2001. *Educational assessment of students.* Third Edition. Upper Saddle River, New Jersey: Merrill Prentice Hall.

Pearson D, Vyas S, Sensale L & Kim Y 2001. Making our way through the assessment and accountability maze: Where do we go now? *The Clearing House*, March/April, 175-82.

Ramsden P 2003. *Learning to teach in higher education*. Second Edition. London: Routledge Falmer.

Schön D 1987. *Educating the reflective practitioner.* Josey-Bass.

Shay SB 2003. The assessment of final year projects: A study in academic professional judgment. Unpublished PhD dissertation. University of Cape Town.

Shay SB (2004). Assessment of complex tasks: A socially-situated interpretive act. *Harvard Educational Review*, **74**(3), 307-329.

Warren Piper DW 1994. *Are professors professional? The Organization of University Examinations.* Higher Education Policy Series, vol. 25. London: Jessica Kingsley Publishers.

Weedon P, Winter J & Broadfoot P 2002. *Assessment: What's in it for schools?* London: Routledge Falmer.

Appendix 1: Extract from the Management and Assessment Planning Protocol (MAPP)

Stage	Description of current practice	Suggested change
E. Supervising E1. Are there explicit guidelines for supervisors? E2. If there are guidelines, what do they cover? E3. What level of help do supervisors give their students? E4. What preparation is given to (new) supervisors? a) project planning skills b) assessment skills c) general communication skills d) none – based on own experience of being supervised e) other. E5. How is the supervision load managed? E6. Is there use of co-supervision? If so under what circumstances?		
F. Assessing F2. What products are assessed? a) written report b) final working model c) written drafts d) oral presentation e) poster f) other. F3. Is the process assessed? If so, how is the process assessed? F4. Are there clear assessment criteria for any of the products or processes that are assessed? F5. If so, how are these assessment criteria communicated to the students? F6. Are there any formative assessments elements (where feedback is given to the student)? F7. If so are these formative elements marked? F8. Who does the assessment? a) supervisor? b) other internal markers (marking or moderating?) c) external markers (marking or moderating?) d) other. F9. Is the marker required to give evidence in support of the mark? If so, how and to whom? F10. What attempts are made to establish inter-marker reliability in the marking? a) consensus on criteria established before marking b) clear guidelines for markers published c) agreement on a marking template (rubric) d) agreement on benchmarks (pass/fail, first class etc…) e) double internal marking of projects f) other. F11. In the event of disagreement, with whom does final authority for the assessment reside?		