

CHAPTER 3

Research questions posed in this thesis.⁵

The work presented in this thesis addresses four major themes regarding the implementation of recent assessment practice advances in medical training programmes in South Africa, a developing country in sub-Saharan Africa. The key themes addressed are: (1) the use of assessment to measure clinical competence, (2) the use of assessment to facilitate student learning, (3) the use of assessment to initiate and sustain curriculum change, and (4) the rational selection of assessment tools on the basis of their utility or fitness for purpose.

The six questions that address the key themes mentioned were primarily selected on the basis of their educational importance, as defined in the literature. In addition, each of these issues has been prioritised, by both the national Department of Health and the Health Professions Council of South Africa, as key assessment practice advances that need to be implemented in all medical training programmes offered in South Africa. Each question is addressed in a separate research paper presented as a chapter in the thesis.

The first paper evaluates the psychometric adequacy and educational impact of a resource-efficient portfolio interview assessment strategy used for the summative assessment of professional competence. In the second paper I explore the use of multivariate generalizability theory to estimate the reliability of a multi-component (composite) examination and predict the most appropriate changes needed to further improve examination reliability. The educational impact of a formative assessment method, based on the observed performance of trainees and the provision of structured feedback, forms the focus of the third paper presented in this thesis. The fourth paper evaluates the observed procedural skills competence of medical graduates commencing clinical practice and the implications regarding training programme relevance. The use of student performance data to endorse implementation of programme innovations, such as problem-based learning (PBL), is the focus of attention of the fifth paper. The final paper proposes a model to facilitate the use of assessment utility parameters in order to select assessment methods appropriate to both purpose and available resources.

⁵ In this chapter references made to the literature are not cited again. They are all cited in the first two chapters of this thesis.

1. Are portfolio interviews a reliable measure of professional competence?

In medical education practice the term “portfolio” has come to mean a “collection of evidence that learning has taken place.” This innovative educational method has gained considerable popularity in health professional training programmes in the past decade. In its simplest format, a portfolio comprises a paper-based collection of evidence indicating participation in a number of professionally authentic tasks usually focused on patient encounters in the clinical workplace setting. Indeed, it is the professional authenticity of portfolios, both in terms of task and location, which largely accounts for their widespread use in health care professional training programmes.

A number of concerns regarding portfolio assessment have, however, been expressed. Firstly, current portfolio assessment strategies require examination times, up to 170 minutes per candidate, which prohibit the use of this innovative learning tool in resource-constrained settings. Secondly, the psychometric rigour of current portfolio assessment methods requires improvement. This may be achieved by clarifying the purpose of portfolio assessment (what is being measured), the use of structured assessment criteria (how it is being measured), standardisation of portfolio entries (the comparability of work assessed) and examiner training (better inter-examiner consistency). Thirdly, concern has been expressed that careful review of the portfolio document, the current standard method of assessment, may not provide real insight into trainees’ clinical ability but simply show that they are good at writing about what they do. More appropriate ways of determining the ability of trainees to deal with complex professional tasks, requiring integration of the relevant cognitive, psychomotor and affective skills routinely used in clinical practice, are needed. The suggested use of a “professional conversation between learner and assessor”, an interview, has not been fully explored in the literature.

In 2002, UCT launched an extensively revised MBChB programme. Despite the assessment challenges associated with the use of portfolios, this learning tool was implemented in the new programme. Fourth year medical students were required to write up a set of case notes reflecting 25 real patient encounters undertaken during a 14-week Internal Medicine clerkship. The portfolio of case notes was examined during a 30-minute interview conducted by a single examiner. Four cases, randomly selected by the examiner, were discussed using structured questions aimed at determining the ability of candidates to interpret and synthesise clinical data gathered during the recorded patient encounters. Examiners were trained to score responses using a global rating scale. The paper presented in Chapter 4 describes an evaluation of the psychometric adequacy and educational impact of this novel summative assessment strategy.

The key purpose of this paper was to determine the psychometric adequacy of a more efficient, in terms of time and human resources, portfolio assessment method that could facilitate and encourage the use of this popular, professionally authentic learning strategy in resource-poor countries. In addition, we were interested in determining the impact of this assessment strategy on student learning behaviour. This paper describes a psychometrically acceptable portfolio assessment tool that measures clinical competence in a manner achievable within the limitations of resource-constrained settings. The wider application of this strategy within the university and the broader national medical education community is mentioned in the closing chapter of the thesis.

2. Do specialist certification examinations reliably measure competence?

Postgraduate specialist certification and licensure is the responsibility of a large number of medical colleges, professional boards and other associations throughout the world. Given the high stakes nature of these examinations, certification bodies have a social and professional responsibility to ensure that they are robust, fair and defensible. To date there is a paucity of published data about the reliability of specialist certification examinations and objective methods for improvement. This suggests a reluctance on the part of institutions to subject themselves to external scrutiny and indicates the ongoing focus on the psychometric evaluation of individual assessment instruments rather than composite assessment packages required to adequately assess the multiple dimensions of professional competence. Psychometric evaluation of these high stakes composite examinations is needed in order to improve current assessment practices and sustain the international credibility of specialist certification processes.

The College of Physicians of South Africa recently initiated a process of review of their composite Fellowship examination. This included the use of generalizability and multivariate generalizability theory to estimate the component and composite reliability of the examination in order to determine its current psychometric adequacy and identify strategies for further improvement. The results are presented and discussed in Chapter 5.

The key purpose of this paper was to determine the psychometric adequacy of a composite specialist certification examination and identify strategies for further improvement. This paper highlights the utility of generalizability and multivariate generalizability theory in determining the psychometric defensibility of high stakes examinations, the professional and social responsibility of postgraduate certification bodies worldwide. Furthermore, it demonstrates the valuable use of prediction studies to explore resource-appropriate ways of further improving these examinations. The impact of this study within the Colleges of Medicine of South Africa is highlighted in the closing chapter of the thesis.

3. Does formative assessment promote learning in clinical clerkships?

Clinical clerkships form the backbone of undergraduate medical education. Unfortunately these apprenticeship attachments are often of limited learning value because students may not engage in sufficient problem solving activities to develop the expected level of clinical reasoning expertise. This may be because students usually encounter patients who have already been assessed by the attending staff. One strategy that may provide students with better authentic problem solving activities in clinical clerkships is the use of “blinded” patient encounters in which students interview and examine patients without access to their clinical records. In addition, feedback, the critical element of formative assessment, may also enhance clerkship learning by informing students of their progress, advising them regarding learning needs, and motivating them to engage in appropriate learning activities. While the rationale for feedback is clear, evidence that the desired outcomes are achieved in the clinical clerkship context is lacking.

Review of undergraduate clinical clerkships at UCT found that limited feedback regarding observed performance often resulted in academically weak students being unaware of their limited competence and usually not seeking assistance during clerkship attachments. It was also found that students mostly engaged in patient encounters after consulting patient records. These findings prompted the implementation of a formative assessment strategy providing structured feedback, based on directly observed “blinded” patient encounters, integrated into bedside teaching sessions conducted during clinical clerkships. Students completing a 14-week Internal Medicine clerkship in their fourth year of study were required to engage in directly observed “blinded” patient encounters during weekly small group bedside teaching sessions. Clinician-educators assessed student performance during these “blinded” encounters and provided feedback using standardised performance rating scales. The paper presented in Chapter 6 describes faculty and student perceptions of the utility (feasibility, acceptability and validity) and educational value of this formative assessment strategy.

The key purpose of this paper was to determine the perceived feasibility, acceptability and educational value of a workplace-based, directly observed formative assessment strategy providing structured feedback. The importance of providing student feedback based on directly observed performance is repeatedly highlighted in the literature but remarkably difficult to achieve, even in well resourced settings. The challenge addressed in this paper was to achieve an acceptable (to clinician-educators and students), feasible strategy with a meaningful impact on student learning. This paper describes a well-liked, resource-efficient formative assessment strategy, based on observed performance, which facilitates student learning in clinical clerkships. The wider application of this strategy within the university and in the broader national medical education community is mentioned in the closing chapter of the thesis.

4. Are the procedural skill competencies of medical graduates adequate?

The ability to perform a wide range of diagnostic and therapeutic procedures competently is a core learning outcome of modern undergraduate medical curricula. Data on newly-qualified medical graduates at the time of entry into their pre-registration (internship) year, however, suggest that the procedural skills competence of interns is frequently deficient when objectively tested. While the need to acquire these basic patient care skills is widely endorsed, educational guidelines often provide only broad educational objectives that do not stipulate the specific practical skills medical students need to master prior to graduation. This has the consequence of producing graduates with widely variable competencies in basic patient care at commencement of clinical practice. Some training institutions have developed detailed undergraduate medical training outcomes, including specified procedural skills, in an attempt to remedy this situation.

The Health Professions Council of South Africa recently published guidelines outlining the educational objectives of undergraduate medical programmes. These guidelines, however, provide broad educational outcomes for accreditation purposes only, and are, therefore, unlikely to promote a uniform level of technical competence in the provision of basic patient care amongst newly-qualified graduates. Guidelines detailing the procedural skills competencies expected of South African medical students upon graduation do not currently exist. To determine the need for such guidelines, we undertook a seven-station OSCE assessment of the procedural skills competence of a cohort of South African medical graduates on commencement of their internship. Candidates' performance was scored using item checklists and the opinion of interns, of their own competence, was sought by questionnaire. The results are presented and discussed in Chapter 7.

The key purpose of this paper was to objectively determine the procedural skills proficiency of medical graduates emerging from South African universities. This information was gathered so as to put forward a cogent argument for the formulation of a national set of basic procedural skills proficiencies expected of junior doctors commencing their internship service in South African hospitals. The paper demonstrates the key role assessment can play in identifying important curriculum changes needed in an environment where curriculum relevance (concordance between programme outcomes and professional practice demands) is at a premium. The impact of this study, both within the university and at a national level, is outlined in the closing chapter of this thesis.

5. Does PBL impact on the performance of academically-at-risk medical students?

Since 1995, race-based distribution of educational resources has been abolished in South Africa. Significant differences remain, however, and the majority of black South Africans continue to be educated in suboptimal circumstances. The reasons for this ongoing inequality are twofold: (1) black children, comprising 93% of South Africa's school-going population of almost 14 million children, cannot all be accommodated in schools previously reserved for white children (less than 10% of school-goers), and (2) the upgrading of services, particularly the provision of adequately trained teachers at previously black schools, will take many years to complete. It is thus not surprising that when black students, emerging from these educationally disadvantaged circumstances, enter university they are at risk of performing poorly despite their best efforts and innate talents.

Attempts to address the educational needs of these students have included the introduction of extended medical programmes at several South African universities. Such a programme, the Academic Development Programme (ADP), was implemented at the University of Cape Town (UCT) in 1991. Over the past decade the ADP has graduated more than 100 students.

In 2002 UCT introduced problem-based learning (PBL) as a method of instruction in the first three years of their revised MBChB programme. Given the success of the ADP, concern was expressed regarding the potentially adverse impact of PBL on the performance of academically-at-risk students. Similar concerns were expressed at the University of New Mexico School of Medicine (UNMSOM) when PBL was introduced there in the late 1970s. In contrast to the decision taken at UNMSOM, we elected to admit all students, including academically-at-risk students, directly to the new PBL programme and carefully monitor their progress. A parallel programme of academic support was implemented for students demonstrating poor academic performance by the end of the first semester. An interim analysis was performed to compare the retention rates and academic performance of academically-at-risk students in the new PBL programme and the ADP. The results are presented and discussed in Chapter 8.

The key purpose of this paper was to determine the impact of a new method of instruction (PBL) on the retention rates and overall academic performance of academically-at-risk students. The paper highlights the value of using student performance data to evaluate the educational benefit of an expensive curriculum innovation in order to provide evidence to sustain change achieved in a setting where both human and financial resource constraints profoundly impact upon prevailing educational practice.

6. How can assessment utility parameters be used to select assessment methods?

Selecting suitable assessment methods remains a daunting task for African medical schools where attempts to revamp existing assessment methods have been hampered by serious resource constraints. Therefore, in developing world regions like sub-Saharan Africa, student performance measures are largely determined by the resources required and the overall feasibility of the method selected. Indeed, these factors probably account for most of the inter-institutional variation observed in the choice of assessment methods used. Furthermore, assessment methods used by most African medical institutions have been empirically selected. This situation, possibly the consequence of difficulties in objectively evaluating factors that influence test selection, highlights the need for a framework to guide the rational selection of assessment methods, particularly in African countries where medical schools are transforming their learning curricula and, therefore, their assessment methods. A lack of guidance on the selection of assessment methods may result in institutions adopting testing methods that cannot be sustained.

The paucity of assessment literature appropriate to the resource-constrained African context, further compounded by adverse socioeconomic, fiscal and political factors, underscores the need for an objective way of selecting resource-appropriate assessment tools. Based on our experience as clinician-educators working in Africa, we formulated a model for selecting assessment methods suitable for use in resource-limited environments. The approach, based on a model used to make rational drug selection(s) in the management of human disease, focuses on four factors influencing test selection: performance, cost, suitability and safety.

The key purpose of this paper was to develop a model suitable for guiding African institutions in the selection of assessment methods appropriate to both educational purpose and context-specific resource constraints. The model, and examples to illustrate its use, are presented and discussed in Chapter 9. The paper emphasises the need to recognise important assessment utility determinants and provides an objective way of accounting for them when selecting assessment tools in resource-limited settings. The potential use of this model in other developing world regions is highlighted in the closing chapter of the thesis.

Concluding remarks

The six questions posed in this chapter are systematically addressed in Chapters 4-9 of this thesis. Each paper focuses on one of the key assessment themes outlined in the literature review in Chapter 1. The socio-political and economic context, within which this work is located, is described in Chapter 2 of this dissertation.