Medical education:
the challenge of linking theory to practice

J.N. Hudson

A thesis submitted in fulfillment of the requirements for the
degree of Doctor of Philosophy (by prior publication)

November, 2004
Abstract

In the 1990s, despite empirical support for linking theory to practice in professional education, a theory-practice gap persisted in medical education. This dissertation presents three initiatives and their evaluation to address the theory-practice gap in medical education at an Australian medical school.

Case Based Teaching (CBT), a teaching and learning initiative, was designed and implemented in the Department of Physiology at the University of Adelaide. It aimed to introduce students to clinical skills in a way that built student confidence; to achieve integration of knowledge between medical and clinical sciences; and to strengthen the link between theory and clinical practice. Students and their tutors reported that CBT was beneficial in their development, and realised its three major aims. Greatest support came from students on the verge of their clinical training.

Few studies, with both internal and external validity, have demonstrated the potential of computer aided learning (CAL) in medical education. In the second initiative, a self-directed CAL resource linking theory to practice was designed in three different learning formats. A randomised controlled trial showed learning gains for CAL users (in terms of ability to retain and apply knowledge) compared to non-users, but the active CAL format did not prove superior to more passive ones. The study demonstrated the importance of using valid assessment tools to measure learning outcomes, and the difficulties
of conducting randomised controlled trials in the real world of medical education.

Finally, the utility of a novel integrated practical examination was evaluated, in terms of reliability, validity, acceptability, cost and educational impact. The large investment of time and effort that was required to reform assessment practice brought rewards. Good reliability was achieved, and validity was maximised by linking theory to practice, and an extensive item review process. Assessment reform was used to drive the integration of teaching and learning.

Basic science and professional knowledge consist of two different worlds, the former being analytic and general and the latter, holistic and particular. This dissertation has addressed the challenge of linking these two worlds, with initiatives that embedded these two knowledge structures within the other, fostering professional competence that is founded on procedural, propositional and personal knowledge.

Key words
Medical education; undergraduate; teaching and learning; computer-aided learning; assessment; integration; theory-practice link; evaluation
Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

J.N. Hudson (Candidate)

Signed……………………………………………………Date……………….

.
Acknowledgements

In addition to previous acknowledgement of the contribution made by colleagues and students to the initiatives in each of the published papers, I wish to thank and acknowledge the following people for their support of this dissertation:

- Dr. Pat Buckley and Professor Caroline McMillen from the Discipline of Physiology, School of Molecular and Biomedical Sciences at the University of Adelaide, for their supervision and intellectual, administrative and personal encouragement to complete the PhD by publication

- Professor Nicky Britten from the Institute of Clinical Education at the Peninsula Medical School, Exeter, United Kingdom for her encouragement and intellectual support, particularly in relation to qualitative research methods

- Dr. Charlotte Rees from the Institute of Clinical Education at the Peninsula Medical School in the United Kingdom for her encouragement, intellectual discussions on medical education, and ideas for the thesis structure
• Dr. Keith Postlethwaite from the School of Education and Life Long Learning at the University of Exeter in the United Kingdom, for intellectual discussion on the theory-practice link in medical education, and ideas for further exploration of the use of science in the medical reasoning process

• My co-authors for some of the studies, Dr. Anne Tonkin, Co-director of the Medical Education Unit at the University of Adelaide, Dr. Pat Buckley and Professor Caroline McMillen (as above), for their permission to submit the published papers to fulfill the requirements of this dissertation

• Professors John Tooke and John Bligh from the Peninsula Medical School in the United Kingdom, for support and encouragement to complete the thesis while I was a senior academic at Peninsula Medical School from 2002-2004

• Dan McHolm for generous IT support during my appointment at the Department of Physiology, University of Adelaide, and in the final stages of my dissertation

• My husband Geoff and children, Sarah, Simon, Christina and Anna, for their continued support of my career and studies
Preface

The three initiatives described in this thesis were designed, implemented and submitted for publication while Nicky Hudson, the doctoral candidate, was employed as a Lecturer in Physiology in the then Faculty of Science at the University of Adelaide, from 1995 to 2002. After relocating to the United Kingdom in April 2002 to take up a position as a Senior Lecturer in Clinical Education (Academic Lead for Physiology) in the new Peninsula Medical School in the south west of England, Nicky was appointed as an Affiliate Lecturer in the Discipline of Physiology, School of Molecular and Biomedical Sciences at the University of Adelaide. The dissertation was completed under the auspices of the latter academic appointment, and at the time of submission of the thesis in mid-November, 2004, Nicky had returned to the University of Adelaide to take up a position as a Senior Lecturer in the Medical Education Unit in the Faculty of Health Sciences.

The initiatives presented in the dissertation have been published as below:


Table of contents

Chapter 1: Literature review

1.1. Introduction .........................................................................................1
1.2. Role of medical education, and expected outcomes .........................2
1.3. History of medical education and professional practice .................9
1.4. Context, values and theoretical underpinnings of Case Based
Teaching ..................................................................................................23
1.5. Context, values and theoretical underpinnings of the Computer
Aided Learning Initiative ..........................................................................34
1.6. Context, values and theoretical underpinnings of the
Integrated Assessment Initiative ..............................................................41
1.7. Summary and implications for the published studies .....................56
1.8. References ..........................................................................................60

Chapter 2: Rationale for the methodology used

2.1. Introduction .........................................................................................71
2.2. Evaluation ..........................................................................................72
2.3. Quantitative and qualitative research methods: theory and practice....73
2.4. Methods used in the evaluative studies in this thesis ......................81
2.5. Ethics ..................................................................................................94
2.6. References ..........................................................................................96
Chapter 3: Linking cardiovascular theory to practice in an undergraduate medical curriculum

3.1. Statement of authorship……………………………………………….101
3.2. Title page for published manuscript…………………………………102

Chapter 4: An evaluation of Case-Based Teaching: evidence for continuing benefit and realization of aims

4.1. Statement of authorship……………………………………………….112
4.2. Title page for published manuscript……………………………………113

Chapter 5: Computer-aided learning in the real world of medical education: does the quality of interaction with the computer affect student learning?

5.1. Statement of authorship……………………………………………….122
5.2. Title page for published manuscript……………………………………123

Chapter 6: Evaluating the impact of moving from discipline-based to integrated assessment

6.1. Statement of authorship……………………………………………….133
6.2. Title page for published manuscript……………………………………134

Chapter 7: Discussion

7.1. Introduction………………………………………………………………147
7.2. Summary, key findings and their contribution to new knowledge…..147
7.3. Methodological limitations………………………………………………160
7.4. Future directions…………………………………………………………166
7.5. References………………………………………………………………170
Chapter 1

Literature Review

1.1. Introduction

This chapter describes the theoretical background, values and context of the published papers contained in this thesis, in relation to teaching and learning, and assessment practice in undergraduate medical education. Section 1.1 outlines the structure of the chapter. Section 1.2 provides a discussion of the role and expected outcomes of undergraduate medical education and section 1.3 describes some important developments in the history of medical education over the past one hundred years, relating them to models of professional practice in vogue at the time. This section highlights the professional model that led to the separation of theory from practice in medical education and the ‘pre-clinical/clinical curriculum divide’, and some empirical support for a more seamless curriculum, where theory is linked to practice from the start of medical education. The following sections, 1.4, 1.5 and 1.6 each focus on the context, values and theoretical background for the published initiatives, namely Case Based Teaching (CBT), Computer-aided Learning (CAL), and the Integrated Assessment, respectively. Section 1.7 presents a summary, which provides a rationale for the work contained in this thesis, and concludes with a presentation
of the research aims for each of the studies. Finally, Section 1.8 lists the references informing the discussion in Chapter 1.

As the dissertation is based on four published articles, that were written to be read as independent works, inevitably there may be some repetition of information between this literature review and the introductory sections to subsequent published manuscripts (Chapters 3-6).

1.2. Role and expected outcomes of medical education

The role of medical education is to develop students’ medical expertise, moving them along a continuum from layperson to novice physician (Patel et al, 1997), competent to deliver the standard of practice and care expected by the public, as well as regulatory bodies such as the Australian Medical Council (AMC) in Australia, the General Medical Council (GMC) in the United Kingdom and the Association of American Medical Colleges (AAMC) in the United States of America. According to *The New Doctor* (2003), the latest publication from the GMC………*

*patients must be able to trust doctors with their lives and well-being. To justify that trust, we as a profession have a duty to maintain a good standard of practice and care, and to show respect for human life.* Their earlier publication, *Tomorrows Doctors* (1993), together with that AAMC publication, *Physicians for the Twenty-first Century* (1992), has played a prominent role in guiding medical educational reform internationally. *Tomorrows Doctors* defined
good medical practice, summarising the twelve broad attributes of an independent practitioner. While these include consultation with others, team-based practice, professional attitudes to practice and education, consideration of ethical issues, health promotion and clinical governance, foremost in the list are the willingness and ability to resolve medical practice problems and understand and advance the knowledge base. In addition to this perspective from the profession itself, medical educators must consider the external or societal expectations of good medical practice when choosing the best approach to educating doctors.

Society views medicine as a profession, the latter being defined by Cogan (as cited in Maudsley and Strivens, 2000b, p.535) as \ldots\ldots\ldots\textit{a vocation...founded upon an understanding of the theoretical structure of some department of learning or science, and upon the abilities accompanying such understanding...applied to the vital practical affairs of man. The practices...are modified by knowledge of a generalized nature and by the accumulated wisdom and experience of mankind. The profession...considers its first ethical imperative to be altruistic to the client.}\n
Many social theorists have outlined the various behaviours that make up professional practice, with each theorist prioritising a different characteristic (Maudsley and Strivens, 2000b, Bines and Watson, 1992). However, Goode (as cited in Maudsley and Strivens, 2000b, p.536) identified two characteristics that are core, claiming all other characteristics are derived from these. Those thought to be core to the definition
of professional practice are........a collectivity of service orientation and a prolonged specialized training in a body of abstract knowledge. It is evident from both internal and external perspectives, that good professional practice is founded on an understanding of a knowledge base.

However, several authors (Eraut, 1995, Schön, 1983, Cogan, 1953) have warned against focusing on the term ‘knowledge base’ in professional education, as it tends to suggest conventional academic knowledge. Eraut (1995) preferred the term ‘professional knowledge’, which includes propositional (‘knowing that’) knowledge, process (‘knowing how’) knowledge, personal knowledge and moral principles. According to Eraut, this better represents the type of knowledge that is used to perform quality professional work. Cogan (1953) also believed that professionals needed to apply more than a conventionally acknowledged knowledge base. As novel or complex problems can arise in practice and are not always solved with research-based theory, he preferred to describe professional practices as being refined by science and corrected by wisdom. For the profession of medicine, this may be equated to the view that there is both an art and science to clinical practice (Schön, 1983, Hudson, 2000). While acknowledging that practitioners of a particular profession share a common body of explicit, more or less systematically organised professional knowledge, Schön (1983) highlighted the complementary role of professional artistry in situations of uncertainty, instability, uniqueness and conflict. For Schön, a key component of competency in professional practice was the capacity for reflection, and this
was central to a model he put forward in 1983, and refined in 1987. He proposed practitioners should be educated to think what they are doing while they are doing it, as well as reflecting after the event (Schön, 1987). This fits with Cogan’s view of practice being corrected by wisdom. Theorising on what constitutes the ‘knowledge base’ for professional practice in general, is particularly pertinent for the education and assessment of students for the profession of medicine.

In medicine, as in other professions, practice is mostly structured in repetitive occurrences of a unit of activity (Bines and Watson, 1992)—in this instance the case or patient visit—and to be competent practitioners doctors, like other professionals, need to have ‘a skill’ based on theoretical knowledge. A doctor’s ‘skill’ or competency is multifaceted, consisting of the knowledge, attitudes and skills that form the basis of clinical practice. Doctors, like other professionals, have recognised tools of the trade, for example the stethoscope. They share a medical vocabulary and a setting, which mostly includes the hospital or doctor’s surgery. Practice, comprising ‘look, feel and listen’, generally involves the doctor and their patient solving a problem, or a concern, that initiates the consultation. The encounter usually consists of getting to know a patient in their psychosocial context; listening to their concerns or symptoms and placing these in the context of their personal and family health, or illness; questioning re further symptoms referable to each of the body systems; reviewing all this information and prioritising diagnostic probabilities; examining the patient with
particular emphases suggested by the history, re-prioritising diagnoses according
to any signs elicited; testing these diagnoses using suitable investigations and
sharing the outcomes with the patient. Obviously, not every consultation in
medicine involves all these steps. For a patient well known to the doctor, data
gathering can concentrate on the presenting episode, in acute trauma the process
will rapidly proceed to the physical examination while for a psychosocial
concern, investigations may not be required. However, a constant feature of
every patient encounter is the personal, interpersonal and clinical skills that the
doctor needs to elicit the symptoms and signs, and the knowledge base that
underpins the reasoning process to solve the problem at hand. To maintain a
good standard of practice and care, a medical practitioner must have an
understanding of what Cogan (1953) termed, ‘the theoretical structure of some
department of learning or science’. While this includes an understanding of the
impact of human behaviour and social influences on health, a doctor’s learning
is founded on an understanding of the structure of the human body, and how it
functions. The development of this foundation is an important role for medical
education.

What about the outcomes of medical education? In vocations like medicine, the
concept of a guaranteed competence is important from the point of view of
providing education to intending and continuing professionals. As Bines and
Watson (1992, p.3) claim........*the student’s qualifications must mean
something in terms of his or her ability to perform certain tasks to a requisite
standard. Undergraduate medical education should take the novice to a level of competency that enables the graduate to carry out supervised, but effective and defensible medical practice. Margetson’s view of the development of competence is useful in the context of this discussion. Margetson (1999, p.363) described learning to become a thoroughly competent practitioner as........the development of an integrated, coherent ‘growing web’ of understanding, knowledge and skill in practice. This ‘web’, which encompasses both a technical command of the knowledge base and the professional artistry that Schön (1983) described, provides the foundation for professional judgement, the process of which is outlined above. As experience is accrued, the web can continue to grow in an incremental way, allowing the novice to become more expert (Maudsley and Strivens, 2000a).

From their work in the field of artificial intelligence, Dreyfus and Dreyfus (1986) described an expert as a practitioner who had an intuitive grasp of situations based on deep tacit understanding, and only used analytic approaches in novel situations and when problems occurred. These findings have been confirmed in the context of medical problem solving. Patel et al (1997) showed that when difficult or unfamiliar problems are encountered, expert doctors have the ability to solve and manage them. It is the explanation of the problem-solving process, critical when problems outside the normal domain of expertise are encountered, that underlies superior clinical competence. Patel and colleagues (1997) showed that ‘experts’ have schema that do more than facilitate
pattern recognition, differentiating them from ‘practitioners’, who solve problems by pattern recognition of common presentations. In her article on the early acquisition of cognitive skill in medicine, Hmelo (1998a) summarises a variety of studies on the characteristics of medical experts. Experts are generally more accurate than novices in their problem solving; they represent their knowledge more coherently; and as their memory organisation is based on specific expertise, they use data-driven reasoning on routine problems. When faced with unfamiliar or atypical problems they use hypothesis-driven causal reasoning. Novices, however, tend to use the latter approach for all problem solving encounters as their memory organization appeared to be based on little or no expertise. Taylor (1994) has also contributed to our understanding of the cognitive structures of novices and experts with his Novex Analysis approach on the development of expertise in other professions.

Schmidt et al (1990) elaborated on the stages of development of medical expertise. They hypothesised that in the journey through to expertise, medical students progress through several transitory stages. These are characterised by distinctly different knowledge representations, underlying performance. As students develop expertise, these ‘older’ representations don’t become inert, but can be activated, when needed. These researchers divide the development of medical expertise into two stages: firstly elaborated causal networks, which explain the causes and consequences of disease in terms of the underlying causal mechanisms, are formed; when exposed to real patients, students develop their
reasoning skills by relating patients’ signs and symptoms to the networks laid down in stage 1. As they see more patients and accrue experience, they start to take shortcuts, only activating the knowledge that is relevant to understand and focus on the case at hand. This model of the development of medical expertise, informed by Schmidt as well as others in the field (reviewed by Hmelo, 1998a), has implications for the education of medical undergraduates. Early experiences should develop hypothesis-driven causal reasoning skills, founded on a basis of general pathophysiological processes, and give students opportunities to gain experience at solving clinical problems. The latter should reflect the real world of the medical practitioner, laying the framework for the ‘web’, which integrates knowledge, skills and understanding. Experience is required to move the novice along in his/her journey to competency in practice (Maudsley and Strivens, 2000a). The challenge is to provide the experience in the contemporary climate of medical education. Before presenting the published initiatives that were developed to support undergraduate medical students in their journey to competence, it is necessary to briefly review the history of medical education to understand the values that underpinned the innovations, and the context in which they were situated.

1.3. History of medical education and professional practice

Western medicine was founded on a base of Hellenic medicine, with Hippocrates’ theory of vis medicatrix naturae, the healing power of nature,
being particularly influential (http://www2.sjsu.edu/depts/Museum/hippoc.html). Social change and advances in scientific knowledge since Hippocrates’ time have had consequences for medical education. It has undergone various transitions, in response to the prevailing view of medical theory, the educational philosophy that is favoured, and more recently, demands for high quality medical care. A review of the transformation of North American medical education over the past 200 years illustrates these points and leads us to the ‘revolution’ in medical education that took place in North America around 1910, with important implications for this dissertation.

The first medical school in North America was founded in 1765 by Benjamin Franklin and by the late 1800s, there were about 400 schools, most of which had no relationship to institutions of higher learning (Papa and Harasym, 1999). Around the middle of the 19th century in the United States, as in England, the original home of the American colonists, professional medicine was influenced by theories of disease derived from the classical Hippocratic texts (Starr, 1982). Illness was seen as a general disturbance caused by an imbalance of the four humours (blood, phlegm, yellow and black bile). After the Newtonian revolution in physics, new medical theories attributed disease to disturbances in solid entities in the body, but the basic model stayed the same (ibid). At the time, two other spheres of practice were equally important as professional medicine, namely domestic, and lay medicine. As the family was the centre of
social and economic life in early America, the woman of the house administered
domestic medicine in the home. Lay or popular medicine, delivered by healers,
occupied the ground between domestic and professional medicine. As Starr
reports (1982), all manner of people took up medicine, for example clergymen,
who combined it with their pastoral role. There was little formal training or
standards of care.

Up to this time, the ‘apprenticeship’ style of learning, originating with the Greek
philosopher Socrates, had been in widespread use. The Socratic method was a
‘dialectic’ or ‘question and answer’ style of learning at the feet of a master. It
was, as Socrates confessed, unscientific. Others have described it as
unsystematic (http://www.2020site.org/socrates/method/html). This style of
professional education, called the ‘pre-technocratic’ model by Bines and Watson
(1992), was used in many other professions at the time. It tended to consist of
on-the-job training, and novices learned by imitating the skilled actions
modelled by experienced practitioners. Competence, consisting of talents, skills
and abilities, rather than knowledge or understanding, was thought to come only
through practice (Bines and Watson, 1992). Bruner (1996) described the
apprenticeship model as one where knowledge just grew as habits, and there was
little distinction between procedural knowledge (knowing how) and
propositional knowledge (knowing that). Practitioners developed the
curriculum, and accountability was employer-dominated rather than an
educational institution having control over the content and character of training (Bines and Watson, 1992).

However, in North America in the 18th century, the roots of professional authority for the medical profession were also developing (Ludmerer, 1985). Standards of knowledge and competency, legitimacy and self-regulation were being examined. Medical societies and medical licensure were introduced during the mid-18th and 19th century, the latter also seeing an increase in the educational requirements for medicine. By 1893, entrance to the John Hopkins University School of Medicine required a college degree (Starr, 1982, Ludmerer, 1985). At that time, exciting findings in scientific research were changing the face of clinical practice. The laboratory offered medical practitioners new ways of diagnosing and combating disease, but placed new demands on the conditions needed for both initial and professional training (Flexner, 1910). In 1910, in response to dissatisfaction with the medical practices taught in many medical schools (e.g. blood-letting and heavy metal poisoning), the apprenticeship philosophy of medical training, and the lack of unity and purpose amongst educational institutions, Flexner (1910) reviewed and reported on all schools in North America. His report was a landmark in the history of medical education. It found that there was an overproduction of uneducated and ill-trained medical practitioners, attributed mainly to the fact that a very large number of commercial schools encouraged unprepared students to undertake the study of medicine. As many medical schools were operating for profit, their income
limited the quality of the instruction they could offer, especially in terms of the facilities needed to incorporate new scientific knowledge. Flexner believed that there would be greater regard for public interests and welfare, if a smaller number of better equipped and conducted schools offered a higher standard of medical education to a smaller number of students. Consequently, medical schools not based on the scientific method were forced to close their doors. By the 1920s, Flexner urged that the practice of medicine should have a firm scientific basis, recommending that the basic sciences should form a fundamental part of medical curricula. His new philosophy had profound structural implications for medical schools.

Over the next forty to fifty years, science assumed a prominent place in medical undergraduate training in the majority of Western medical schools. In most of the teacher-centred learning environments that were subsequently adopted, ‘scientific experts’ imparted their knowledge to learners in the first two to three years of the course. This first phase, where students received a discipline-based grounding in theory, became known as the ‘pre-clinical’ years. It was followed by two to three years of practical training, the later phase being termed the ‘clinical’ years. The curriculum structure that Flexner’s reform imposed was in keeping with the ‘technocratic’ model of professional education (Bines and Watson, 1992). As was the case in medicine, professional schools became associated with institutes of higher education, and the institutions tended to have control over the character and content of the education and training. Educational
courses were taught by academic subject specialists, or by practitioners with the appropriate academic credentials. These educators specialised in interpreting and applying selected academic insights for the development of a theoretical base for practice, directing and assessing their own discipline-based courses (ibid). The academic departments of the institution delivered this Phase 1 academic education, which was followed by a period of supervised practice in the workplace. The latter, a major component at the end of the course, was mostly delivered by the professional school. Despite the discipline separation, some courses had varying degrees and types of integration. Content consisted of a systematic knowledge base, which was interpreted and applied to practice. It was multidisciplinary, and included values, as well as knowledge (ibid).

It was the ‘technocratic’ model that ushered in the ‘pre-clinical/clinical’ divide. Theory was no longer learned, and thus stored in the brain, in the context of practice. Students were left to make the link between theory and practice in the later years of their course. This created the challenge for medical education, a challenge that is ongoing, and the theme of this thesis: to link theory to practice. According to Bruner (1996) applying theoretical knowledge to practical problems has ‘troubled thoughtful people forever’, exemplified by the attempts to apply science to medicine. He reports that even Aristotle was troubled by this problem, and quotes from Aristotle’s book, the Nichomachena Ethics (Book V, 1137a)………It is an easy matter to know the effects of honey, wine, hellebore, cauter, and cutting. But to know how, for whom, and when we should apply
these remedies is no less an undertaking than being a physician. Even with the scientific advances, Bruner (1996, p.44) states that………*the challenge is always to situate our knowledge in the living context that poses, the “presenting problem”, to borrow a bit of medical jargon.*

In the 1950s, an ‘organ-system-based’ model was introduced to reduce the amount of disjointed, discipline-specific information in the curriculum, and efforts were made to ensure the remaining basic science was co-ordinated and relevant (Papa and Harasym, 1999). Although this resulted in the integration of basic science content around body systems, the information was still disjointed for learners. The latter did not learn how to use the organ-system theoretical framework to solve a clinical problem, a primary cognitive task for a clinician (*ibid*).

Gradually over the latter half of the 20th century, medical educators increasingly became troubled by the problem of applying theoretical knowledge to practice. In Canada, late in the 1960s, one response to the theory practice-gap was Problem-Based Learning (PBL), a new learning process for medical education. PBL was valued as an active, collaborative learning environment, which was student-, rather than teacher-centred (Galey, 1998, Hmelo, 1998b). It aimed to integrate the basic, human and clinical sciences, in the context of a virtual case. By exploring self-generated learning issues, as the symptoms, signs and investigation findings associated with the ‘problem’ were progressively revealed, the student
generated his/her own learning agenda. Supporters claimed that this format should develop enduring and meaningful neural networks, where theoretical knowledge would be linked to, and thus stored, in the context of clinical practice (Hmelo, 1998a). Over the next 30 years, the PBL process was adopted in many medical schools, often as part of the SPICES model proposed by Harden and colleagues in 1984. This model proposed student-centred approaches to learning, problem-based learning, integrated teaching, a community orientation, elective study periods and systematic approaches to curriculum planning (Harden, 1984).

Integration was the important educational strategy central to this wave of reform. Many medical schools attempted some level of integration, as opposed to compartmentalisation of the various scientific disciplines underpinning clinical practice. Harden (2000) has summarised the various levels of integration, with discipline-based teaching at one end, and full integration at the other. At the top step of his integration ladder, called ‘Transdisciplinary’ learning (synonym—fusion, immersion, authentic), the curriculum transcended the individual disciplines. Harden (2000, p.555) states that........the focus for transdisciplinary integration for learning, however, is not a theme or topic selected for this purpose, but the field of knowledge as exemplified in the real world. He cites as examples the final phase (2 years) of the medical curriculum at Dundee, Scotland and the integrated clerkships in ambulatory-care settings at South Dakota, USA. However, it is noteworthy that these examples of high-level integration are in the final phases of the course, achieved after phase 1 has taken
place. The curriculum remained non-seamless, marked by the pre-clinical—clinical divide.

All medical curricula since the early 1990s have acknowledged the important contribution of science to medical practice. However, the wave of reform that started in the 1960s, led educators to revisit the role of science in undergraduate medical curricula. There were further calls to reduce the science content, make it more relevant and build bridges between the scientific knowledge base for medicine and the professional knowledge which clinicians use in practice (reviewed in Maudsley and Strivens, 2000a). While in 1988 Calman and Downie were still arguing for the intrinsic value of teaching basic science to all prospective medical practitioners, concerns were gradually being raised about the ability of, or the ease at which, practising clinicians could apply their basic science knowledge to clinical situations.

Researchers began to examine the role of science in medical education. In a study where students from a traditional (technocratic) curriculum were asked to relate clinical case vignettes to neuroanatomy, Balla (1980) exposed student difficulty in applying basic science. Unlike neurologists, who were able to describe ‘pictures’ of the nervous system triggered by the clinical case, students couldn’t make the appropriate connections. Patel et al (1988), in one of a series of studies investigating the role of biomedical knowledge in clinical reasoning, showed that medical students couldn’t use basic science information to explain a
patient problem, if the basic science was given to them prior to the clinical problem solving experience. Others noted that many medical students couldn’t apply their knowledge to practical medical problems, although they were able to answer test questions (Galey, 1998). Many students memorised the relevant basic concepts in biomedical sciences, without actually understanding them (Regan-Smith et al, 1994).

In follow up work, again in a traditional medical curriculum, Patel and co-workers (1990) found that when responding to clinical case descriptions, students’ pathophysiological responses were incorrect. While these students use of basic science knowledge varied with increasing clinical experience, it remained inconsistent (Patel et al, 1990). They concluded that when the basic science information was introduced out of the context in which it was to be used, students couldn’t make the necessary link between the information and its clinical context. According to Patel and colleagues (1990), the reason both students and clinicians from these curricula had difficulty in relating basic science to clinical cases, was not because basic science information lacked utility for the clinical reasoning process, but it was due to the fact that basic science and clinical context are different knowledge structures. They explained that there were two ways of knowing: medical science was said to be analytic and general, while professional knowledge was holistic and particular (as cited in Boreham, 2000, p.505), and there was no natural embedding of the knowledge structures of the distinct worlds of basic science and clinical medicine within the other (Patel et al, 1990).
was just some of the work that started to question the traditional emphasis placed on medical science in undergraduate curricula, and it prompted further work in understanding how the two worlds of theory and practice are related.

With another group of investigators, Patel studied the performance of final year medical students, house staff and specialist endocrinologists in a patient workup, investigating whether the structure of the doctor’s pathophysiological explanation revealed connections with the prior data (Patel et al., 1997). They found that an understanding of the underlying causal mechanisms is necessary for effective patient management and therapy. They suggested that medical student explanations are inaccurate and limited in scope, as their pattern of reasoning is based on a lack of adequate domain-specific knowledge to generate a relevant schema. Together with Schmidt et al. (1990), they described their analysis of the student problem solving process. Students start to develop their reasoning skills by relating patients’ signs and symptoms to elaborate causal networks laid down in the first stage. As their experience with patients grows, students become more efficient, using only the knowledge that is relevant to solve the case at hand. This provided support for developing students’ hypothesis-driven causal reasoning skills, founded on a basis of general pathophysiological processes, and providing opportunities to gain experience at solving clinical problems. Research in assessment by researchers such as Elstein and colleagues (1978) also showed how problem solving was closely linked to knowledge. The ‘content-specificity’ of problem solving, i.e. the ability to solve
problems is dependent on the clinical content of the problem, will be described further in section 1.6.

Science was acknowledged for its important role underpinning problem solving. Problem solving ability is a core skill for doctors, not only for effective practice but also to reduce the cost of unnecessary tests, irrelevant surgery or extraneous medication. Clinicians must also be able to explain science to patients in everyday practice, e.g. when counselling patients on issues of diagnosis, and therapeutic and non-therapeutic management. An explanation of the mechanism of action, and side effects, of a particular medication, can help doctor and patient (now more knowledgeable, thanks to information technology) reach a shared agreement, about its use, facilitating adherence to therapy. Science still had an important role, termed ‘imagination’ by Bligh (2003), to stimulate intellectual curiosity in those who would go on to advance the ‘knowledge base’. However, the fundamental issue of integration remained. Patients and their problems don’t fall into neat categories. As Bruner (1996) explained, the challenge is always to situate the knowledge in the living context that poses the ‘presenting problem’.

The concern about the fragmentation of overall learning, into discrete and unrelated parts, including the disjunction between theory and practice, was not just confined to medicine. As a result of his research into education in a variety of professions, Schön (1983) started expressing dissatisfaction with the model he described as ‘technical rationality’, as it failed to……..reflect the nature of professional knowledge and action and the way in which professionals actually
develop their practice (Bines and Watson, 1992, p.13). Subsequently a third model of professional education, the ‘post-technocratic’, emerged during the 1980s (Bines and Watson, 1992). The model placed emphasis on the acquisition of professional competence, using a view of professionalism and professional education, which reflected the nature of professional knowledge and action, and the ways in which professionals actually developed their practice. According to Schön’s theory of practice, the knowledge and techniques of technical rationality (science and research) are “bounded and mediated by the arts of problem-framing, implementation and improvisation” (Bines and Watson, p.16). He suggested that a ‘practicum’ was a better environment in which to prepare for professional practice. Here guided by skilled practitioners or coaches, learners could develop through experience of, and reflection on, practice (Schön, 1987). The practicum, situated in the academic institution or employment-based (or both), potentially was a bridge between these two settings. According to Bines and Watson (1992) adoption of the practicum requires an understanding of the range of competencies involved in professional practice, how students acquire professional competence, how skilled teaching could facilitate this, and the best setting for learning and assessment.

Some contemporary medical education courses have features of the ‘post-technocratic’ model. The PBL environment may be seen as a ‘virtual practicum’. It includes knowledge for practice, and there is also an emphasis on the research skills, that Schön (1987) claimed were needed to use reflection, observation,
analysis and evaluation, to develop practice. However, PBL only integrates the theory: the basic and clinical sciences. The practical aspects tend to stand alone in clinical skills training. There has been a recent trend of introducing clinical skills into the ‘pre-clinical’ curriculum. In many schools, particularly in Europe, skills training starts in purpose built clinical skills laboratories (Bradley and Bligh, 1999). These are ‘protected’ environments where students practise and receive feedback on their skill acquisition, at a time when service pressures in the hospital and shorter patient stays have reduced access to clinical tutors and patients, respectively. They do however tend to emphasise performance of skills, and mostly fail to combine this practical experience with explanations of why it is done. True integration, that is continually connecting practice and theory (Bruner, 1996), is not achieved.

By the 1990s, due to the historical and empirical perspectives described above, three goals were being espoused for undergraduate medical education: to integrate basic medical and clinical science with professional competencies through the five years, to be supported by integrated assessment; to reduce the burden of factual information by providing a core curriculum (as well as options); and to encourage self-directed learning supported by informatics resources (Tomorrows Doctors, 1993). The two learning environments and novel assessment, published as part of this thesis, were a response to both these international, and domestic, drivers for change. The following three sections describe the context, values and theory underpinning them.
1.4. Context, values and theoretical underpinnings of Case Based Teaching

Context and Values of Case Based Teaching (CBT)

CBT was designed in the mid 1990s as part of curriculum reform at the University of Adelaide in South Australia, and in response to student requests that their basic sciences teaching be more relevant. At this time, guidelines such as *Tomorrow’s Doctors* (1993) were driving the revision of medical curricula all over the world. The Report of the Working Party on Medical Education (*British Medical Association*, 1995), endorsed recommendations for self-directed learning, and argued for learning environments, which featured collaboration not competition, and links between theory and practical experience. Thus CBT was a response to both local and international imperatives for change.

With CBT, there was a move away from what Berstein (as cited in Bines and Watson, 1992, p.19) has described as *collection code*, with strong boundaries between disciplines, towards what they described as *integrated code*. This underpinned the initiatives introduced in the year 2 Human Structure and Function course in Adelaide. As anatomy and physiology underpin the practice of medicine, they are core, compulsory elements of any medical course. They share common boundaries with other medical sciences, offering exciting
opportunities for the integration of theory, centred on medical cases (as described in Chapter 3). However, the integration was taken one step further. CBT linked the theory not only to the interpretation of the symptoms, signs and investigation findings, but also to their elicitation by students. According to Bruner (1996), this represents ‘true integration’, meaning the learner closely combines practical experience with conceptual explanations of why it is done. CBT was a simulated practice setting, a version of the practicum described by Schön (1983), situated in an educational institution. To develop student confidence, it fostered observation, modelling, routine practice, problem solving and reflection, partnering students with experienced practitioner facilitators and their peers. Flexner’s philosophy that the practice of medicine should have a firm scientific basis was still valued, the educational process was case-based learning in small groups, but its novel feature was the linking of knowledge to professional practice. By adopting this curriculum structure, we aimed to give learners the foundation for a ‘seamless’ development towards competence.

As Bines and Watson (1992) state, it is not easy to integrate theory and practice, especially when the practicum is based in the educational institution. This was indeed the case. The contributing disciplines had concerns about the elements of CBT, which developed professional knowledge and understanding rather than expertise in the discrete disciplines. CBT put pressure on the time that was left for the relevant disciplines (e.g. traditional anatomy or physiology practicals), and required a different kind of tutor to facilitate the sessions. Clinical
practitioners, with knowledge of the other undergraduate-course elements, and ability and interest in linking them, were required. By situating the practicum in the academic institution, the course co-ordinator (a practitioner and an academic staff member of the University) could structure the learning environment and experiences to support the aims of the students and the organisation. By using medical tutors who modelled appropriate attitudes and values, students started their socialisation into the medical professional culture and identity.

**Theoretical basis of CBT**

Amongst the calls for reform from the GMC in 1993 (*Tomorrows Doctors*) was a recommendation that modern educational theory should inform medical education practice. The expression of CBT in undergraduate medical education was informed by several educational theories. Firstly it drew on the theoretical framework ‘andragogy’, a term first introduced by Knowles, who defined it as the ‘the art and science of helping adults to learn’ (Kaufman *et al*, 2000, p.4). It has roots in humanistic psychology, and according to Kaufman and colleagues, the basis of ‘andragogy’ is that........*the attainment of adulthood is marked by adults coming to view themselves as self-directed individuals* (ibid). In response to criticism that this model was not a theory but a set of assumptions, Knowles altered his model to describe ‘andragogy’ as part of a continuum from ‘pedagogy’ to be used according to the time and situation, and not just depending on the learner’s age. Current thinking on this model is that
‘andragogy’ captures the general characteristics of adult learners, and does offer guidelines for planning instruction with learners who tend to be at least somewhat independent and self-directed (ibid). Knowle’s conceptualisation of ‘andragogy’ (as cited in Maudsley and Strivens, 2000b, p.542) is illustrated in Table 1.

Table 1. Knowles conceptualisation of ‘andragogy’

<table>
<thead>
<tr>
<th>Adult education involves learners:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Knowing why they need to know</td>
</tr>
<tr>
<td>• Being internally motivated</td>
</tr>
<tr>
<td>• Being ready to learn</td>
</tr>
<tr>
<td>• Taking responsibility for their own learning</td>
</tr>
<tr>
<td>• Using their own experience</td>
</tr>
<tr>
<td>• Using life-centred, task-centred, problem-centred approaches</td>
</tr>
</tbody>
</table>

This conceptualisation of andragogy underpinned both the case-based, and self-directed, learning approach of CBT. However, CBT also introduced new experiences, including the first step of physical contact with the ‘patient’. Tutors were pivotal in establishing a safe setting so students were willing ‘to have a go’, and not fear being humiliated in front of their peers. They role modelled the appropriate way to give feedback on student-derived hypotheses,
facilitated discussions, and encouraged students to take part as both ‘doctors’ and ‘patients’ in history-taking and physical examination activities. Students assuming these roles when comfortable and therefore relaxed, were thought to feel less vulnerable to failure and more likely to succeed than physiologically stressed learners. The first of seven principles that Knowles drew from the assumptions of ‘andragogy’ to guide adult learning activities was pertinent for the CBT culture. The principle stated that an effective learning climate should be established. Learners should be comfortable, both physically and emotionally. They should feel ‘safe’ and unpressured to express themselves without judgement or ridicule (Kaufmann et al, 2000. p.5).

So adult learning theory was applied to both behavioural and cognitive aspects of CBT, with the virtual case serving to integrate the theoretical learning in the basic and clinical sciences (See Chapter 4, Aim 3). However, it didn’t fully encompass how the new clinical skills were introduced and learned. They were acquired not only by learning through direct experience, but also by observing the actions of tutors and peers. The ‘modelling’ aspect of CBT was grounded in aspects of Social Cognitive Theory, proposed by Bandura some twenty to thirty years ago (Bandura, 1986). His theory also combined behaviourist and cognitive approaches to learning, highlighting their social or interactive aspects (Kaufman et al, 2000). The basic tenet of Bandura’s social cognitive theory is that our actions, learning and functions are the result of continuous, reciprocal, and dynamic interaction between personal, environmental and behavioural
determinants, the relative influence of each depending on the activities and the circumstances (*ibid*). This is illustrated by comparing two environments in the medical education continuum. Students bring their own personal determinants to each, but in CBT students were asked to attend to the task of learning new clinical skills, and in this context the behavioural determinant was able to exert more influence. However, in the busy ward, because of the service pressure in the hospital, the environment is influential and medical students have to adjust their behaviour to ‘get the job done’ (Kaufman *et al*, 2000). Bandura viewed humans as possessing five capabilities that underpin our learning and functioning in all situations. One of these, our ability to learn vicariously as well as experientially, is thought to shorten the learning process (Kaufman *et al*, 2000). CBT offered students both experiential and vicarious learning experiences as they interacted, observing and modelling the behaviours of their tutor and/or peers. The outcomes for each student depended on their individual needs, values and goals (personal determinants).

Bandura (1986) also emphasised the importance of cognition in mediating our actions, claiming that people’s judgement of whether they are capable of dealing with a task, their self-efficacy, determines whether they will choose to participate, how much time and effort they will invest in the face of disappointment and whether they approach it anxiously or confidently. This was the theoretical basis for another important aim of the CBT initiative, namely to introduce students to clinical skills, in a manner that built confidence for their
first ‘real’ patient encounters (See Chapter 4, Aim 1). ‘Confident’ describes…...a judgement which influences whether an individual is willing or not to undertake an activity (Stewart et al, 2000, p.903). The literature contains reports of students who have felt afraid when asked to perform clinical tasks that they have been neither taught, nor practiced (Stewart et al, 2000). This can have a negative effect on their self-confidence, making them more likely to avoid these tasks later, and contributing to a negative self-efficacy about the task.

CBT, and especially its major aim, to strengthen the link between the theory and practice of medicine (see Chapter 4, Aim 2), drew from the tenets of ‘situated learning’ theory, and the ‘reflective practitioner’. Situated learning theory questions the value of knowledge transmitted by instruction (Brown et al, 1989), claiming that ‘learning to do’ (closely related to ‘knowing how’) takes place through solving problems in context, stressing the importance of other people who are experienced at solving similar problems, to the context (as cited in Maudsley and Strivens, 2000b, p.537). Brown et al (1989) are critical of education practices that separate what is learned, from how it is learned and used, arguing that knowledge is situated, being in part a product of the activity, context, and culture in which it is developed. One example used to illustrate their claims is Miller and Gildea’s work on vocabulary teaching (ibid). They compared learning language from dictionary definitions and a few exemplary sentences, to the way vocabulary is normally learned outside school, and showed
that……..learning from dictionaries, like any method that tries to teach abstract concepts independently of authentic situations, overlooks the way understanding is developed through continued, situated use (ibid, p.33). According to Brown and colleagues (1989), knowledge is like language in this regard: a concept, like the meaning of a word, is always under construction, and part of its meaning is always inherited from the context of use. They make a useful analogy between conceptual knowledge and a set of tools. Tools, like knowledge, can only be fully understood through use. Use involves changing the user’s view of the world, and adopting the belief system of the culture in which they are used. They claim……..it is possible to acquire a tool, but to be unable to use it, just as it is possible for students to acquire algorithms, routines, and decontextualised definitions that they can’t use (Brown et al, 1989, p.33). A relevant example may be a medical student who is full of book learning, and knows about the use of a stethoscope, but is unable to use these tools to solve the problem of someone who is short of breath.

Brown and colleagues (1989) applied situated learning theory to support a new model for learning, termed ‘cognitive apprenticeships’. Activity and social interaction were seen as key to ‘enculturating’ students into authentic practices. ‘Apprenticeship’ emphasises the centrality of activity in learning and knowledge, and highlights the context-dependent and situated nature of learning (Brown et al, 1989). They also chose the term as it suggested the paradigm of……..situated modelling, coaching, and fading (ibid, p.39), the process that
teachers use to promote learning in such a setting. After making their tacit knowledge explicit, or modelling their strategies using authentic activity, teachers support students’ attempts at doing the task. The final step is to empower students to continue independently (Brown et al, 1989). This learning environment was called a ‘cognitive’ apprenticeship as it offers more than the usual apprenticeship techniques, including the kinds of cognitive skills more normally associated with conventional schooling. Brown et al (1989) state that this is not as incompatible with traditional apprenticeship as it may at first seem, saying that……..The physical skills usually associated with apprenticeship embody important cognitive skills, if our argument for the inseparability of knowing and doing is correct (ibid). When the technocratic model of professional education replaced the apprenticeship model it separated knowing from doing. CBT, informed by situated learning theory and the ‘cognitive apprenticeship’ deriving from this, aimed to link them again.

Schön’s theory of the ‘reflective practitioner’ was the final major influence on the development of CBT. Reflection is a key issue in professional behaviour and learning, and according to Kaufman et al (2000), the concept of the ‘reflective practitioner’ is the essence of the epistemology of professional practice. Schön introduced this theory in the 1980s (Schön 1983, Schön 1987). He argued that the ‘knowledge-first, application-later’ approach to professional education was in danger of producing practitioners full of theoretical knowledge that they were unable to use and apply to practical situations. His new
epistemology of practice was not only based on reflection-on-action or thinking after solving problems (Schön, 1983) but also on reflection-in-action or thinking while problem solving (Schön, 1987). According to Schön (1987), an important part of the competence and artistry in practice of skilful practitioners was their ability to think what they are doing while they are doing it. His model of the reflective practitioner drew from several epistemologies: positivism, interpretive theory and critical theory. However it is of the positivist view of science, the framework for the ‘technical rationality’ model of professional education, that Schön was most critical. As Kaufman and co-authors (2000, p.15) explain…...the positivist view of science assumes that educational theory is a scholarly pursuit unrelated to practice. Rather than seeing only the predictive view of theory as of practical value, Schön claimed that theory and practice inform each other, and as knowledge is embedded in practice, practitioners are well placed to test and revise theories by reflection and action (ibid). According to Kaufman and colleagues (2000), the reflective process serves as a bridge in the theory-practice relationship, tying the ‘reflective practitioner’ to interpretative theory, an approach that proposes that theory is interpreted in the light of personal current and past experiences. Finally, Schön’s reflective practitioner model draws from critical theory, the tenet that theory is intimately linked to practice through a process of critical thinking and examination. Kaufmann et al (2000) then provide a succinct summary…...as a result of this process professionals are then free to break free from established paradigms and reformulate the ways in which practice, problems, and problem solving are
viewed...This re-framing is part of learning and change...It is how practice helps organise theory. Schön argued that in schools where knowledge was separated from practice, there was no opportunity to develop reflection-in-action. This created a dilemma of rigour or relevance for educators, practitioners and students (Bines and Watson, 1992).

To combine the teaching of the ‘applied science’ with coaching in the artistry of reflection-in-action, Schön put forward his concept of the practicum (already mentioned above), referred to by Bines and Watson (1992) as a virtual world.........intermediary between the lay world of ordinary life, and the esoteric world of the profession, or a setting mid-way between self-directed learning, i.e. learning on your own, and apprenticeship learning, i.e. being attached to a senior practitioner in the real world of professional practice (ibid, p.16). As described above in relation to Bandura’s social cognitive theory, in a practicum such as CBT one has time for both educational aims, and rudimentary and useful practice skills, skills that practitioners expect of students who are apprenticed to them later in busy hospital ward. Clearly a practicum, based on the theory of the reflective practitioner, shares features and values with a cognitive apprenticeship, based on situated learning theory. Both provided important elements of the theory on which CBT practice was founded.
1.5. Context, values and theoretical underpinnings of the Computer Aided Learning (CAL) Initiative

The context and values of CAL initiatives

Computer-aided learning may seem a long way from Aristotle, and the physician who must know how, for whom and when remedies, such as hellebore and cautery, should be used. Thanks to modern science and technology we have an ever-increasing understanding of human structure and function, and sophisticated techniques to investigate health and disease. The medical practitioner must keep abreast of this knowledge, while still needing to know how, for whom and when to use it. Thus in contemporary practice, more than ever before, self-directed learning has become an important element in the development and maintenance of professional competence. Computers and electronic learning (e-learning) have, and will continue to, open up new opportunities for learning in undergraduate education and ongoing training. Some go as far as claiming that... adopting new approaches to the provision of education and training, including the utilisation of new learning technologies is not an option—it is an imperative (Inglis, as cited in Harden, 2002, p.469).
Computers clearly offer exciting potential for improving student learning, either as an aid or replacement to traditional formats, or for the development of innovative approaches. There are many reasons to fund computer assisted learning, namely: convenience and flexibility; unique presentation benefits; personalised learning for users; economies of scale; competitive advantage to schools using it; the expansion of pedagogical horizons; and achievement of the ultimate goal of higher education, to link people into learning communities (Greenalgh, 2001). In his editorial on myths and e-learning, Harden (2002) advises caution. He discusses the ‘hype’ around e-learning, and possible social consequences of adopting it as a learning environment. However, Harden (ibid) believes the reality is that… e-learning is playing and will continue to play an increasing role in medical education. In doing so, it will not replace face-to-face learning but will be part of a blended curriculum.

Sophisticated simulation models demonstrated the power and potential of CAL in medical education back in the 1960s (Harden, 2002), but this was not realised at the time due to technical limitations. Dramatic advances since then for example, the development of personal computers and the Internet, have allowed technology to have considerable impact on education in terms of teaching, learning and assessment. The past three to four decades have seen an ever-increasing number of CAL programs incorporated into health science curricula, and on-line learning has served as a stimulus for the development of new instructional approaches. The CAL program described in Chapter 5 is what has
been termed, a reusable learning object (RLO) (Harden and Hart, 2002). Some have claimed that learning objects are the technology of choice for the next generation of instructional design, development and delivery (Harden, 2002). RLOs have the advantage of allowing students to progress at their own pace, but educators need to be aware that issues such as gender, culture and student perceptions can impact on the ease at which CAL is embraced. For example, reports in the literature have shown that despite evidence to the contrary, students did not believe that computer instruction was time effective to enhance learning outcomes in medical physiology (Richardson, 1997).

While in the CAL evaluative study (Chapter 5) students accessed their assigned tutorial individually, CAL need not be ‘lonely learning’ (Harden, 2002). RLOs can also be used as a collaborative learning environment by groups of students, a format to be encouraged, as evidence suggests that students may learn better from each other than from experts (Peile, 2004). While learning from experts has been the traditional foundation of learning in medicine, Peile (2004) tendered evidence that experts are not always the best people to teach as they have become unconscious of the processes that novices need to learn. In the initiative which is the focus of Chapter 5 the tutorial designer, with enthusiasm for the topic and an intermediate level of proficiency in neuroradiology and the neuroscience underpinning it, was potentially more sensitive to the needs of the student learners, than the expert who kindly reviewed the tutorial and its content.
Computer presentation is well suited to topics such as radiology, which are visually intensive. In the radiology tutorial, computer presentations replaced the use of a specialist radiologist (of high resource cost) giving repetitive traditional tutorials to small student groups, year after year. While students enjoy the interaction with specialist tutors, the reality was that due to high service demands and waning University budgets a ‘reusable’ CAL tutorial was more practical. If positive learning outcomes could be demonstrated, it would be an enduring, cost-effective and student-centred learning resource. Thus the CAL initiative was developed in the framework of a University imperative to develop e-learning initiatives, in line with current global trends in education generally. It was only a few years before a proposal for an International Virtual Medical School (IVIMEDS), a blend of the best of e-learning and face-to-face learning (Harden and Hart, 2002). While we needed to respond to the concerns of students and teachers, and myths about e-learning, which can become barriers to engagement, the reality was that it was playing a major role in medical education, a role that was likely to increase. The debate was on how best to use it, grounded in the principles of best educational practice, and how to respond to the needs of individual students in terms of support and learning format.

The CAL initiative (Chapter 5) was designed in the late 1990s, at a time when the six-year medical undergraduate curriculum was a hybrid of both traditional and problem-based learning approaches. Neuroanatomy and neurophysiology were traditionally introduced to medical students at the University of Adelaide in
the third year of their undergraduate curriculum, in a co-ordinated course on structure and function of the nervous system. The Department of Physiology implemented CBT sessions to introduce neurological clinical skills, and aid students to integrate their developing understanding of neuroscience with clinical neurology (Hudson, 2004, submitted to *Medical Teacher*). The CAL tutorial on basic neuroradiology was a self-directed learning resource so students could prepare for CBT sessions. Investigations like imaging provide new information for the reasoning process. They are ordered after reflection on the symptoms and signs, with test findings either confirming hypotheses or leading to new actions, the type of professional activity described by Schön (1983). The signs and symptoms in the CBT neurological case promoted thought about diagnostic possibilities, which were refuted or confirmed by student interpretation of the CT and MRIs, using their e-learning derived knowledge.

The same values espoused for the CBT learning initiatives hold true for the CAL approach. The tutorial was well planned with identified objectives, opportunities for repetitive use and feedback, and assessment, which was congruent with the curriculum goals. Again the challenge was not only to blend the CAL into the curriculum, but as for face-to-face initiatives, to take a holistic approach and link theory to practice. This was achieved by linking the neuroanatomy, neurophysiology, pathology and basics of neuroimaging, to their use in practice. The CAL tutorial was developed in three different formats. In all formats, apart from the images, the information was displayed as text.
However, the didactic version was in the pure-text format commonly used in
textbooks, and a passive experience for the user. While the problem-based
tutorial was a little more interactive, the free-text format was the most active
learning experience, requiring the user to respond to questions and submit free
text answers, for which immediate feedback was given. An active learning
environment has been defined by Modell (1996, p.69) as………one in which
students engage in the process of building and testing their own mental models
from information that they are acquiring. As Glaser (1991) showed that
teaching to develop this type of cognitive activity, rather than teaching to deliver
knowledge, fosters competence, superior learning outcomes were hypothesised
for the active format. This value was tested in the evaluation described in
Chapter 5.

Theoretical basis of the CAL initiative

The CAL initiative was underpinned by both self-directed and adult learning
theory. Self-directed learning is a natural process in human life. In his
introduction to Hammond and Collin’s book on self-directed learning, Boud
states………Learning is about a great deal more than acquiring knowledge and
developing skills. If learners do not also develop the capability of directing their
own learning, and acting on the world around them, they will be only partially
educated, and limited in what they can do (Hammond and Collins, 1991, p.1).
In the educational literature, self-direction has been framed in two different
perspectives. Some theorists have a humanistic orientation and see self-direction as a goal towards which individuals strive, having autonomy, choice and personal responsibility for their learning. This overlaps with the alternative approach developed by other workers in the field who view self-directed learning as a way of organising learning and instruction, and the tasks to varying extents, are in the learner’s control (Kaufman et al, 2000). An early contributor to the second perspective is Knowles, the theorist who introduced the assumptions that define ‘andragogy’. As highlighted earlier, the basis of andragogy is that adulthood is attained when individuals see themselves as self-directed individuals, and that andragogy offered guidelines for planning instruction with learners who tend to be at least somewhat independent and self-directed (ibid). Some of the features of adult learning environments, previously illustrated in Table 1, apply to the self-directed computer learning tutorials.

Self-directed learning is also an aspect of several other theoretical approaches. For example, in Bandura’s social learning theory (1986), two other capabilities that humans are seen as possessing, are the capacities for self-regulation and self-reflection. These allow us to regulate our behaviour according to our own standards and evaluation of our actions, and using metacognition (thinking about our thought processes) to learn from experience (Kaufmann et al, 2000). Many cognitive, humanist and constructivist theorists also incorporate self-directed learning into their framework but in different ways (ibid). For an educator, acknowledging that the capability to self-direct one’s own learning is important,
a computer reusable learning object can offer a setting in which to express this intention. Clearly, self-directed learning is also an important element in the development and maintenance of professional competence, so is a critical skill to develop in would-be medical practitioners.

1.6. Context, values and theoretical underpinnings of the Integrated Assessment Initiative

The context and values of Integrated Assessment

Assessment is a necessary practice in any course, aiming to determine whether students have achieved a standard, which in the case of professional practice, is set by the educational institution and/or the profession employing the graduates. For medicine, both these bodies and the public need to be assured that graduates can practice as safe, responsible, and knowledgeable doctors. One way of looking at assessment in medicine, according to Schuwirth (2004, p.16), is to see it as........a measurement of medical competence, adding that assessments could then be regarded as........diagnostic tools. He warns us of the dangers of false-negative and false-positive outcomes in assessment as they have consequences for the student, the medical school and society. Thus it behoves us to ‘get it right’.
At the time the integrated assessment (Chapter 6) was developed and implemented at the University of Adelaide in the late 1990s, there was generally a more holistic approach to medical education. Teaching and learning were concerned with three main aspects of professional development: the professional knowledge base, competence in professional action and the development of reflection (Bines and Watson, 1992). The interrelationship of these three elements, and the emphasis on competence, led to integrated initiatives such as CBT (Chapter 3). The growth of different modes of learning had implications for assessment, for example in terms of the content, modes and timing of assessment. However, with the emphasis on competence, one needed to take care not to fall in to what has recently been described in the literature as the ‘monkey see, monkey do’ approach (Talbot, 2004).

Everyone would agree that a doctor should be competent, but according to Talbot (2004), a competency model in medical training can limit the reflection, intuition, experience and higher order competence necessary for expert, holistic or well-developed practice. Enthusiasm for a model that sees competency as fulfilment of standards determined in advance could lead to competency, rather than developing what Talbot calls competence i.e. competent professional performance. His concerns are expressed in the context of anaesthetics postgraduate training in the United Kingdom, but are ones also relevant to undergraduate medical education. Competence is necessary in many vocational situations, and Talbot advises that it should be viewed as multi-layered, open and
reflective, and have a focus on dialogue and argument, rather than in a ‘monocultural classification’, as the latter can encourage sameness and conformity. Talbot (2004, p.589) presented three views of competence: operational competence (knows how), academic competence (knows that), and life-world becoming (reflective knowing). These views represent the three models of professional education presented in Section 1.3: the pre-technocratic, the technocratic, and the post-technocratic respectively. Reflective knowing, Talbot’s preferred view of competence, and the one underpinning the CBT and CAL initiatives, values practical understanding. Understanding is an important word in this context, as Talbot (ibid) says........understanding brings with it a critical edge, and in this era of evidence-based practice, a critical edge is a priceless tool for the professional. Competence becomes stuck in an authoritarian certainty, while one’s understanding may change tomorrow: that, surely, is the true nature of professional practice. The higher order model of competence i.e. performance underpinned by understanding, was the model that the assessment initiative aimed to test.

The innovative assessment, an integrated practical examination (IPE), was implemented at a time of curriculum change at the University of Adelaide. The medical curriculum was a hybrid of both problem-based and traditional learning strategies, but still had the ‘flavour’ of discipline-based teaching and a pre-clinical and clinical divide. There was some unease that the previous assessments, a critical hurdle between the pre-clinical and clinical learning
environments, gave the wrong message to students by encouraging discipline-based rather than integrated learning. There was some staff resistance to integration of both teaching and assessment, but students were demanding change. A collaborative reform of the third year assessment presented an opportunity to address the desires and concerns of all, with assessment ‘leading the way’ in terms of integration.

The IPE was part of a new assessment programme. Together with integrated multiple-choice questions (MCQs) and modified-essay questions (MEQs) it replaced six discipline-based written and practical assessments (Hudson and Vernon-Roberts, 2000, Tonkin and Hudson, 2001). MCQs test knowledge, and the ability to apply it, while MEQs assess student knowledge and ability to critically appraise information and solve problems. Both had been shown to have good reliability (Tonkin and Hudson, 2001), the former enabling sampling of a broad domain of knowledge. The IPE was added to the assessment program to test students’ ‘practical understanding’. It replaced an Objective Structured Clinical Examination, which assessed the clinical, practical and communication skills of students, and separate Anatomy, Microbiology and Pathology Structured Practical Exams (sometimes called Spotter Exams), which tested basic science knowledge in these disciplines, particularly in terms of recognition of structures. The novel feature of the multi-station IPE was that stations (items), starting with authentic tasks in clinical practice, sought students’ ability to use an integrated understanding of clinical, basic and human sciences, as does
a competent medical practitioner. The anatomical, pathological and microbiological specimens associated with the tasks were used to foster and test understanding.

The rationale for the development and evaluation of the IPE was informed by both an understanding of historical developments in, as well as the theoretical basis of, assessment practice over the past sixty or so years. As they are inextricably entwined, the theory underpinning the integrated assessment will be presented in association with its relevant historical context.

**Theoretical basis of the integrated assessment**

Historically, advances in assessment practice have followed, often slightly lagging behind, advances in teaching and learning practice. In the era of the pre-technocratic model, assessment tended to be ‘on-the job’. Examinations, mostly in the control of the employer rather than the educational institution, concentrated on the candidate’s skills. With the technocratic model, the knowledge needed for competent practice became key, and institutions assumed a major assessment role (Bines and Watson, 1992). Following the Second World War, the steep rise in the number of students enrolled in higher education posed logistical constraints. This, together with dissatisfaction with assessment practice in the health sciences, prompted the search for new diagnostic tools to assess clinical competence (van der Vleuten, 1996). At that time, there was an
atomistic view of competence (Hager et al., 1994), which was described by the prevailing literature in personality and educational psychology as a ‘trait-conception’. In the Latent Trait Theory, competence was seen as an aggregate of different components or latent attributes, which were seen as relatively distinct from each other and relatively stable across clinical situations and time. Competence grew as each component was developed in learning experiences, in a monotonic way (van der Vleuten, 1996, p.42). This conception of competence led to assessment methods and formats that measured each component separately, and varied according to the target trait.

However, part of the dissatisfaction with assessment practices in the early 1950s was due to approaches that were a legacy of the apprenticeship style of teaching and learning, for example methods such as implicit assessment; holistic judgements; and non-standardised tests (van der Vleuten, 1996). These tended to be subjective and had poor measurement characteristics (ibid). The response to this, together with increasing student numbers and the emphasis on knowledge following Flexner’s reform, was the introduction of MCQs. The expansion of computer technology helped with this task and MCQs were adopted worldwide. While time consuming to construct, MCQs can handle large numbers of exam candidates, and can efficiently and objectively sample many different domains of knowledge achieving excellent reliability (ibid). However, the MCQ soon came under criticism itself. As van der Vleuten (1996) explained, MCQs were accused of measuring trivial knowledge and having a ‘cuing’ effect. Tools that
measured higher levels of cognitive functioning, and authentic professional tasks, were sought.

Written simulations, such as Patient Management Problems (PMP) and Modified Essay Questions (MEQ), were introduced in the 1960s to try and overcome some of the disadvantages of performance tests (such as subjectivity), while still having a high degree of realism or fidelity (Kane, 1992). In these tests, sometimes with the aid of the computer, a realistic patient problem was used to simulate reality and assess clinical reasoning ability and the problem-solving process. They proved popular and were included in many high-stakes examinations. Research on written simulations raised three significant outcomes for assessment (as summarised by van der Vleuten, 1996). Firstly they revealed a phenomenon, which has been called the case-specificity or content-specificity of problem-solving (Elstein et al., 1978). The ability to solve problems was found to be dependent on the clinical content of the problem, as a score arising from one problem was not predictive for a score on another. Therefore, generalisation of the results from the observed performance to a universe of similar observations was a problem unless large numbers of simulations were used (Kane, 1992). Secondly, the hypothesis that competence (according to the trait-conception of competence in vogue at the time) grew with increasing expertise was not confirmed. Experienced clinicians failed to consistently score more than less experienced clinicians or students (van der Vleuten, 1996). Thirdly, problem solving seemed to be much more closely linked to knowledge,
and was not as independent a construct as was originally thought. For example, researchers such as Norcini et al (1985) reported very high correlations between reliable scores on written simulation problem-solving tests and another measure, the MCQ. Apart from theoretical concerns, these findings had resource implications. A large number of test items were needed to achieve acceptable levels of reliability. For example, with a favourable average inter-problem correlation of 0.30, at least 10 problems were needed to obtain an accepted and arbitrarily defined alpha of 0.80 (van der Vleuten, 1996). Written simulations lost favour (Kane, 1992). One response was the ‘key feature approach’ to the assessment of problem solving, where questions consisted of a realistic case description followed by a small number of questions requiring only essential decisions (Schuwirth and van der Vleuten, 2003). Despite good reliability and validity in measuring problem solving ability, they were also costly to produce and haven’t been widely adopted.

As discussed in Section 1.3, educational reform in the 1970s and 1980s encouraged an active approach to teaching and learning. Rote learning of passively acquired knowledge was discouraged, but assessment had not kept pace. It still tended to reward unwanted patterns of learning behaviour. Efforts to develop instruments that measured the learning process produced assessments such as the Triple Jump Exam (van der Vleuten, 1996). While measuring problem-solving skills and students’ ability to gather information, it also proved costly to administer, in terms of staff time, and potentially suffered from the
content-specificity of problem solving. However, learning process measures did highlight the educational value of assessment and how it drives learning (ibid), a concept to be noted when aiming to change students’ learning behaviour via assessment practice.

An important implication for assessment in the 1990s was the introduction of competency-based assessment a decade or so earlier. No matter how technically correct the assessment procedures were, the atomistic approach to standards was thought to be less valid, as were competency standards based on a generic approach, for example just testing whether candidates could use their knowledge. Context was seen as vital to competence (Hager et al, 1994). In 1994, Hager and co-workers defined competency-based assessment as……..the assessment of a person’s competence against prescribed standards of performance (ibid, p.3), and competency was conceptualised in terms of……..knowledge, abilities, skills and attitudes displayed in the context of a carefully chosen set of realistic professional tasks, which are of an appropriate level of generality (ibid, p.4).

In medical education, in vivo, rather than in vitro simulation was adopted (van der Vleuten, 1996, p.47), where the actual performance of exam candidates was assessed in standardised live simulations of clinical situations. Internationally, the prototype most widely adopted has been the Objective Structured Clinical Examination (OSCE), a single test consisting of a number of ‘stations’ (Smee,
2003). At each station, examinees are asked to perform a particular skill or to manage a patient, and a trained examiner assesses student performance using precoded checklists and/or rating scales. Standardisation, maximised by the use of checklists and examiner and patient training, plus the fact that the item tasks approximate the real world, have made the OSCE a popular assessment format. While they demonstrated the same findings as written simulations, i.e. high correlations with other measures, content-specificity of items and sometimes absence of differences between levels of expertise, the OSCE has been adopted as part of the assessment programme of many medical schools (van der Vleuten, 1996). Van der Vleuten (1996) attributes this to the fact that it is standardised professional testing approximating real life.

In 1996, in light of the research findings in modern assessment practice, a simple model was proposed by van der Vleuten (1996) to determine the utility of an assessment. The model highlighted five variables: reliability, validity, educational impact, acceptability and cost, that should be considered and assigned different values or weightings, depending on the user’s situation and needs. Validity and reliability are particularly important when referring to the usefulness of a measure to assess professional competence. If an assessment measure is invalid or unreliable, it is not reasonable to make important assumptions or decisions, based on the values it produces (Schuwirth, 2004). As assessment is a powerful stimulus for learning (van der Vleuten, 1996) it can have considerable educational impact, especially at a time of assessment reform.
Sharing assessment reliability, validity and educational values with all stakeholders was novel for assessment practice when integrated assessment was introduced at the University of Adelaide, and was potentially a practice that could increase the educational impact of assessment. Thus for the IPE evaluation (Chapter 6), reliability, validity and educational impact were assigned most value, and will be discussed in turn.

Reliability

Reliability refers to the consistency with which any particular instrument measures a construct (Judd et al, 1991). If an instrument is reliable it will produce the same result no matter who uses it. In assessment, it is generally used to look at the consistency or reproducibility of examination scores—the extent to which variation in a set of grades represents systematic differences among individuals, rather than other sources of (error) variation (Swanson et al, 1987). So the score that each examinee obtains must be reproducible in a parallel examination of similar difficulty. At the very least, the parallel test should rank students in the same order, i.e. there should be correlation between individual student rankings in each test (Schuwirth, 2004). Reliability, expressed numerically from 0 (no reliability) to 1 (perfect reliability) indicates the degree of correlation. A reliability of 1.00 means an instrument is completely free from error, which is virtually impossible in practice (Judd et al, 1991). Reliability can be measured in several ways, but Cronbach’s alpha
For nearly all methods of assessing professional competence, the number of items is crucial for reliability: tests containing a small number of items whatever the format, produce unstable or unreliable scores (van der Vleuten, 1996). Other sources of variability can impact on examination reliability, for example in an OSCE, the examiner, the patient and the rating method. Training of the former, and standardisation of the latter, for example by the use of checklists make these sources of variability easier to control (van der Vleuten, 1996). Most studies of clinical simulations have used classical test theory to investigate reliability. According to Swanson et al (1987), classical theory encourages the use of a ‘one-source-of-error-variance-at-a-time’ approach to estimating reliability, and often overestimates true reliability. Consequently they prefer to apply generalisability theory to estimate reliability as in reality, scores are simultaneously influenced by all sources of error variance. A generalisability coefficient, ‘g’, which takes into account all sources of variability, is used to measure reliability rather than Chronbach’s alpha. In research into the reliability of role-playing oral examinations as a function of patient-case and examiner sample size, Swanson (1987) using a g coefficient, showed that reliability improved if a different new examiner was used for each case for each examinee, as a candidate’s judgement was then based on many raters, and the bias due to
examiners was averaged out across cases. Such research findings contributed to
the view that for a measure to be reliable, it must gather a sufficient sample of
candidate performance, and score it with some level of standardisation and
structure (Swanson, 1987, van der Vleuten, 1996).

Validity

Validity is less straightforward than the concept of reliability, and is usually only
guessed or estimated by the researcher. It refers to the extent to which the
measure we use is able to ‘tap’ the true construct, or has construct validity (Judd
et al, 1981). There are a number of different kinds of validity, and those relevant
to this dissertation are discussed herein. In assessment, validity refers to whether
the test actually measures what it intends to measure but as it is difficult to
define perfect criteria or standards for the construct of clinical competence,
claims about assessment validity are usually based on a variety of sources
(Schuwirth, 2004). According to Schuwirth (2004), the literature on validity can
be divided into approaches based on the expectations of examinations scores,
and those that look at test content. Firstly, in terms of expectations, experts
should score higher than non-experts in a valid test (a measure of indirect
validity). There is also an expectation that a valid test would correlate well with
one measuring the same characteristics (convergent validity). While the high
correlations found between written and multiple-station tests (van der Vleuten,
1989) did not necessarily mean that the same construct was being measured, it
did challenge the notion that the unique characteristics of the assessment method determine what particular aspects of competence are being measured. Subsequent research has showed that validity more likely depended on what van der Vleuten (1996, p.51) called the *stimulus format* of a test item rather than the *response format* in which the answer is captured.

The notion that what is put in to the method is more important than the method itself, is the basis of the second view of validity, one that suggests looking at test content (Schuwirth, 2004). A test of high content validity must be perceived to contain relevant and meaningful items by all the major stakeholders, and those with mastery of the topic, should be able to answer all items correctly. Further support for content validity is derived from matching of the test topics and/or tasks with clearly defined curriculum goals, a process called ‘blueprinting’ (*ibid*) and by using experts and consensus in the developmental process (Hutchinson *et al*, 2002). Blueprinting ensures that the assessment covers the areas intended, i.e. the learning outcomes given to staff and students. One way of doing this is to produce an assessment blueprint or matrix, with outcomes along one axis and the number of times they have been tested in various assessments, along the other (Hall, 2001).

Consequential validity is a relatively recent, and increasingly important concept, which looks at the effect of assessment on learning, and the political use of test
results. It brings in the ethical aspects of assessment (Hutchinson et al., 2002), and can be seen as part of the educational impact of an assessment.

**Educational impact**

Any assessment action results in an educational reaction (van der Vleuten, 1996). Snyder (1971) writes of the ‘hidden curriculum’ to describe the fact that if the examination does not match the expressed curriculum objectives, and the examination defines academic success, students will respond to the curriculum expressed in the examination. For strategic learners, it is a natural response for academic success (Aaron and Skakun, 1999). The driving force of examinations on learning has important consequences for test developers, so care must be taken to avoid unwanted outcomes.

Acceptability to all the stakeholders and cost are also important issues in assessment. Assessment not accepted by staff and students will not survive even if it has excellent psychometric properties. Provision of information, and addressing staff and student concerns can improve acceptability, enabling assessment reform (van der Vleuten, 1996). There are very few reports in the literature on the real costs of assessment practice. Further research is warranted as good assessment is expensive, and limited budgets can lead to trade-offs between learning and assessment initiatives. However as highlighted above,
assessment costs can be justified in terms of their learning, as well as assessment outcomes.

1.7. Summary, and implications for the published studies

Margetson (1999), like eminent educational theorists such as Schön (1983, 1987) and Bruner (1996), has been critical of educational experiences that separate theory from practice. They oppose medical curricula that consist of a pre-clinical phase, followed by a clinical one, and argue for rigorous, authentic problem-based learning experiences where understanding is integrated with knowledge and skill. It is Bruner who reminds us that true integration is only achieved when a learner continuously connects theory and practice, i.e. the learner should closely combine practical experience with conceptual understandings of why it is done. Reflection on the history of medical education has revealed how ‘practice’, initially the focus of teaching and learning, became secondary to a ‘grounding in theory’ as a result of sweeping curriculum changes in the 1920s. However, by the 1960s, discontent and deficiencies associated with the ‘theory first-practice later’ approach were starting to appear in the literature. In many contemporary medical curricula, problem-based learning was introduced to achieve a level of theoretical integration, and training in clinical skills was introduced early to help students acquire clinical skills. However, this often meant that theory ran parallel to, rather than being integrated with, practice. In the 1990s, and ongoing, the challenge remained for medical
education: to link theory to practice and ensure that knowledge, understanding and skills were woven together early in the journey from novice to competent doctor.

Information technology has provided new opportunities for teaching and learning in medical education (Harden, 2002). Apart from merely delivering information, there has been potential for computer-aided learning to encourage autonomous, independent learning in a wide range of areas. As for face-to-face learning in medicine, electronic learning initiatives must be well planned and evaluated to determine their full educational potential. ‘True integration’, the value espoused for face-to-face learning, should also underpin electronic learning. This also holds true for assessment, as it has been shown to be a powerful drive for learning.

Developments in assessment have shown there is no ‘assessment panacea’: there are strengths and weaknesses for each method. For example, in the interest of reliability, one should use tests that take a broad sample from the domain that is being assessed; performance tests are more likely to reflect the real world of medical practice. When deciding what assessment measures to adopt, it is important to determine the model of competence to which one aspires. As it is difficult to design an assessment that perfectly measures clinical competence, the best option is a programme consisting of a variety of carefully selected and combined assessment methods (Schuwirth, 2004).
In summary, by the mid-to late-1990s in medical education, further effort was needed to narrow the theory-practice gap created some eighty years earlier. New educational settings such as electronic learning had emerged, but their potential was not fully determined. Research evidence suggested that theory prior to practice was not beneficial for medical problem solving, and questioned the usefulness of assessments that measured theory separated from practice. This created the context for the published integrated learning and assessment initiatives. The articles, together with their research aims, are as follows:

- The Case Based Learning (CBT) studies reported in Chapters 3 and 4 describe the CBT innovation and evaluate student and tutor perceptions of its effectiveness. These studies aimed to address whether CBT had achieved its aims, and explore undergraduate medical student and tutor perceptions of the CBT course.

- The Computer Aided Learning (CAL) study reported in Chapter 5 seeks to determine the effectiveness of a self-directed computer tutorial, in terms of the most effective learning format in which to deliver the material. This study aimed to determine whether a CAL tutorial in an active format resulted in superior learning (ability to retain and apply knowledge) to that obtained in more passive formats.
The Integrated Assessment (IPE) study reported in Chapter 6 describes the development and implementation of a novel assessment, and evaluates its utility as judged by students and staff. This study aimed to determine whether the IPE was a useful assessment tool, in terms of its reliability, validity, educational impact, acceptability and cost.
1.8. References


Flexner, A. (1910), Medical Education in the United States and Canada: a report of the Carnegie Foundation for the Advancement of Teaching, Bulletin No. 4, Updyke, Boston, MA

General Medical Council. (2003), The New Doctor: Recommendations on general clinical training, General Medical Council, London


Hippocrates

[http://www2.sjsu.edu/depts/Museum/hippoc.html](http://www2.sjsu.edu/depts/Museum/hippoc.html)


Modell, H.I. (1996), “Preparing students to participate in an active learning environment”, *Advances in Physiology Education*, vol. 270, no. 6, pp.69-77


The Socratic Method and Doctrine


Chapter 2

Rationale for the methodology used

2.1. Introduction

This chapter describes the theoretical and ethical background, and nature, of the methodologies used in the evaluative studies in Chapters 3 to 6. Section 2.1 outlines the sections contained in the chapter. Section 2.2 discusses how evaluation is defined in several ways, according to its purpose or emphasis. Later, in the discussion, each definition will be linked to the pertinent study in the thesis. Section 2.3 presents the theoretical background of the quantitative and qualitative research methods used, together with a debate on the rationale that informed the choice of methodology in the studies. It is followed by section 2.4, which briefly describes the nature and/or features of the methodology relevant to this dissertation; and section 2.5, a summary of the ethical guidelines, and the principles from which they were derived, that underpinned the thesis evaluation practice. The chapter concludes with section 2.6, a list of the references used.
2.2. Evaluation

In education, as in health and other social services, evaluation is an essential part of any intervention or innovation. Evaluation is defined, in the literature, in different ways according to its purpose or emphasis. For example, in the context of health services, St. Leger et al (as cited in Murphy et al, 1998, p.222) define evaluation as........*the critical assessment, on as objective a basis as possible, of the degree to which entire services or their component parts fulfil stated goals.* An alternative definition, stressing measurement and rationality (the mainstream evaluative model), saw evaluation as being........*based mainly on rational thinking and on measurable data. It is not initiated to prove a particular point of view but with the aim of improving a particular activity (ibid).* The following World Health Organisation (WHO) definition however, explicitly and intelligently reconciles all the different approaches (ibid)........*Evaluation is a systematic way of learning from experience and using the lessons learned to improve current activities and promote better planning by careful selection of alternatives for future action. This involves a critical analysis of different aspects of the development and implementation of a programme, its relevance, its formulation, its efficiency and effectiveness, its costs and its acceptance by all parties involved.* In addition, Shortell and Richardson (as cited in Murphy et al, 1998), advise evaluators to acknowledge the political nature of evaluation itself, and not just attend to improved design or more precise measurement. This can be important for an evaluation that is
debated before it is accepted, and implemented, with funding coming from a variety of sources. In such cases, the question to be addressed and the interests the evaluation would serve, are important features. For example in medical education, the question, ‘Is it a useful test?’ has to be studied in broad terms, not only to show all stakeholders its psychometric features, but to also determine its impact on, and acceptance by, staff and students. Assessment not accepted by staff and students will not survive (van der Vleuten, 1996). The term ‘persuasive’ has also been used to describe an evaluation when the results form part of the strategy to support or ‘attack’ a political position or stance (Rossi and Freeman, 1993). An example, again in the context of medical education, is when the results are used to attack a stance such as ‘discipline separation’, in favour of integration.

2.3. Quantitative and qualitative research methods: theory and practice

Bland et al (2000) in their paper on the evaluation of substantive curriculum change in medical education claimed that the choice of evaluation methods should be tailored to the goals of the evaluation process. Traditionally quantitative (‘countable’) methods have been used to evaluate curriculum quality, but Bland and colleagues (2000) remind us that in curriculum evaluation, the data generated by qualitative methods can offer rich descriptions
and explanations of phenomena to complement the quantitative information. Unlike quantitative methods, qualitative research involves the collection, analysis and interpretation of data that can’t easily be reduced to numbers. It is valuable for the exploration of subjective experiences (Carr 1994), and has been used in education research for a considerable period of time (Bogdan and Biklen, 1982). Methods such as focus group interviews and open-ended written questions are qualitative methods that are useful for some of the ‘how’ and ‘why’ questions (Frye et al, 2000). Frye and colleagues showed that when investigating curriculum reform in medical education, a combination of both quantitative and qualitative research methods supplies the means for a multifaceted approach. They advise that when choosing a method to fit the purpose, the researcher should consider the following: the merits of each method, the nature of the research question posed, and knowledge that the findings from both methods can complement each other. The rationale for this advice will be elaborated below, as it is the approach used in the evaluative studies in the dissertation.

There has been considerable debate in the literature on the relative merits of qualitative and quantitative methods, and the appropriate relationship between them. There are those who argue that qualitative and quantitative methods are best seen as deriving from fundamentally different paradigms (Murphy et al, 1998), with each paradigm having a different underlying philosophy and set of assumptions (Dootson, 1995). According to Munhall (1989, as cited in Murphy
et al, 1998), the assumptions should guide the research process by providing a philosophy within which research questions can be developed. While there is still some lack of consensus about the assumptions that underpin both research methods, it has been conventional to characterise quantitative research as being underpinned by a rationalistic paradigm, which stems from the branch of science known as positivism or empiricism (Dootson, 1995). Positivism was concerned with truth and prediction. Followers studied parts to understand the whole, and explained nature through observation, the testing of scientific hypotheses and the development of theories (ibid). In the rationalistic approach the subject is reduced to a reactive, passive being and the environment is controlled in an attempt to seek absolute truths (Dootson, 1995, p.183). Interaction between the researcher and the participants is rare. In contrast the naturalistic paradigm (Dootson, 1995), stems from a more recent view of science known as constructivism, the paradigm on which much qualitative research is grounded (Murphy et al, 1998). Constructivism rejects the notion that the world is composed of facts and that the role of research is to uncover them (ibid). Guba and Lincoln (1989) describe the constructivist position as being grounded on the tenet that there exist multiple, socially constructed realities ungoverned by laws, natural or otherwise...these constructions are devised by individuals as they attempt to make sense of their experiences...constructions can be and usually are shared...this does not make them more real but simply more commonly assented to. The constructivist paradigm recognises that findings
exist due to the interaction between the researcher and the respondents, and takes full advantage of the interaction between them (Murphy et al, 1998).

A more modern view of the two-paradigm approach sees qualitative research being grounded in scientific idealist assumptions, and quantitative, being based on scientific realism. Realists believe that theories do exist; the claims they make are either true or false; and the object of science is to establish the truth about how well the world operates. Idealists, however, view the external world as consisting mainly of representations, and as a creation of the mind (ibid).

Others argue that it is the logic that underlies qualitative and quantitative research that makes them fundamentally different, with the former adopting inductive logic and the latter using deductive procedures. Murphy et al (1998) describe these approaches as follows: deductive research begins with a theoretical system, the concepts of that system are operationalised, and then it sets out to gather empirical data to test that system. In contrast, inductive research starts with the collection of data and moves from there to a general conclusion. Usually a general principle is derived from a set of specific observations. However, this isn’t totally accurate or always useful in practice. While qualitative research has particular strengths in terms of hypothesis generation, not all of it is primarily or exclusively hypothesis generating. Both quantitative and qualitative research can involve induction or deduction at different stages (Murphy et al, 1998).
The two practices are also differentiated on the basis of claims that qualitative research focuses on participants’ meanings and emphasises description. However these categories don’t include all qualitative, and do include some quantitative, research. Rather it is more useful to recognise the complementarity of the two methods, acknowledging that each method has particular strengths and weaknesses. In their excellent review of the relationship between qualitative and quantitative methods which informs the discussion above, Murphy et al (1998) also advise researchers to decide which approach is likely to answer the question at hand in the most effective and efficient manner, rather than deciding on qualitative or quantitative methods on the basis of philosophical or ideological grounds, adding………*Neither can establish truth claims with absolute certainty, but should aim to result in knowledge that is relevant and about which we are reasonably confident* (Murphy et al, 1998, p.86).

Qualitative research has a longstanding use and is widely accepted in education research, but some medical educationalists have claimed that it lacks rigour. Bligh and Anderson (2000) attribute this to a perception that only controlled experimental studies can meet quality criteria, and there are very few of these in medical education research. Traditionally four major criteria have been used to judge the rigour of quantitative research, namely validity, generalisability (external validity), reliability and neutrality (Murphy et al, 1998). Debate continues on the appropriateness of these criteria for assessing the quality of qualitative research, and the details of the various arguments are beyond the
scope of this thesis review. However, some relevant points will be summarised and briefly discussed. Achieving reliability in the traditional sense has been rejected as impossible in qualitative research, as these methods study what is assumed to be in flux, multi-faceted, and highly contextual \( (\text{ibid}) \). Observations can change for reasons other than error, for example the researchers’ insights and sensitivities may change \( (\text{ibid}) \). Lincoln and Guba (1985) suggested that researchers could comment on the dependability, rather than the reliability, of a study if they examined how their research design may have induced change in the inherently unstable phenomenon they were studying.

Vermeire and co-workers (2002) state it is the commitment to rigour, clarity and systematicity that is important, regardless of the method adopted. This principle has been extended to the sampling decisions in qualitative research (Murphy \textit{et al}, 1998). As in quantitative research, pragmatic decisions should be integrated with sampling in a systematic way and opportunistic sampling should be avoided or minimised. In addition to clarity of methodology and systematic data analysis, Stacy and Spencer (2000) suggest that trustworthiness, reflexivity and triangulation are other criteria that can be used to assess the quality of qualitative medical education research. Trustworthiness, or credibility was thought to be more appropriate than conventional validity criteria, such as internal and external validity (Lincoln and Guba, 1985). For example, in a focus group study, credibility refers to the endorsement of the researchers’ conclusions by the subjects of the research (Murphy \textit{et al}, 1998). While there are benefits in
using participant checking to reduce errors, there can be problems associated with this. Apart from the fact that this concept comes from one particular paradigm and the way researchers believe they can understand the world is contentious (Stacy and Spencer, 2000), participant responses can potentially challenge rather than lead to endorsement or validation of the analysis (Murphy et al, 1998).

Quantitative research aims to eliminate the impact of the researcher on the findings, and its traditional criterion of neutrality or objectivity, rests on assumptions of the possibility of freedom from values (Lincoln and Guba, 1985). However, as qualitative research takes full advantage of the interaction between researcher and subjects, it needs to be explicit about the role that the investigator played in the research. Reflexivity is the sensitivity to the ways in which the researcher’s presence in the research setting has contributed to data collected, and how their prior assumptions have shaped the data analysis (Murphy et al, 1998). Thus reports on the outcomes of qualitative research methods need to explain how the researcher’s presence in the setting has contributed to or influenced the data collected, as well as the effect of their epistemological stance on the analysis (Stacy and Spencer, 2000).

Triangulation, a method developed from surveying, is another tool that can be used to assess evidence. In surveying, several points or co-ordinates on a map are more useful in locating an exact position, than one. In research, triangulation
has been defined as a combination of methodologies to study the same phenomenon (Murphy et al, 1998), and four main types have been identified: method, data, investigator and theoretical. In the former two, different methods and data sources are used to study the same phenomenon, while in the latter two, different investigators or theoretical models are used in the same study. It has been advocated to overcome the biases attached to a single method, but some warn that if used inappropriately, triangulation may compound the weaknesses of a research project, rather than strengthen it (ibid). Triangulation has been proposed as a means by which the validity of qualitative research findings can be enhanced or established, as more than one approach gives greater confidence to the findings (Stacy and Spencer, 2000). Some endorse its original use, i.e. to test the validity of the findings of one of the approaches used, aiming to establish the convergent validity of findings from complementary approaches. Others have supported its use in the interests of completeness, to draw attention to the different perspectives that may be revealed from different contexts, aiming for divergent validity (Murphy et al, 1998). Stacy and Spencer (2000) urge researchers to seek explanations for unique variance which otherwise may have been neglected by single methods, rather than using them to dismiss findings from triangulation as flawed.

Many researchers advise that to produce ‘good’ and useful research, we must look at ways of assessing the quality of our evidence; there is no such thing as context-free evidence; we must use judgement to consider the relevance of
research findings to our own settings; and we should critically analyse approaches, using criteria that are appropriate for the paradigm underpinning the research (Harden et al, 1999, Murphy et al, 1998, Stacy and Spencer, 2000). As Murphy and colleagues concluded (1998), for both research methods we should aim for some measure of validity, and relevance.

2.4. Methods used in the evaluative studies in this thesis

Randomised controlled trial

The randomised controlled trial (RCT) is generally acknowledged as the gold standard of evaluative research, the key feature of randomisation being the establishment of equivalent groups at baseline (Torgerson, 2002). It is a research method underpinned by the positivist paradigm. Oakley (1998, p. 1240), in her brief history of experimentation and social interventions, reminds us that the original meaning of control is...check—the word coming from counter-roll, a duplicate register or account made to verify an official account, and that the term ‘control’ entered scientific language in the 1870s in the sense of...a standard of comparison used to check inferences deduced from an experiment. In the early 1900s, the need for a control group in experiments was noted, and was followed in 1908 by the use of the pre-test, post-test design, where the experimental group had an intervention but the control group had nothing in between the tests. However, the idea of randomisation and that of
study ‘blinding’ (i.e. the researcher and/or participants are unaware of the group to which they have been randomly assigned) was introduced into psychology experiments in the 1880s (Oakley, 1998).

In the first half of the 20\textsuperscript{th} century, the advantages of prospective experimental studies with randomly chosen controls was seen as an important solution to linking interventions with outcomes in the social fields, and educational researchers, especially in the United States, went on to lead the way in the design development of RCTs. In education, as in other fields, there was a need to tackle the issue of\ldots\ldots causal inference—how do people know that what they do works, and how can they reasonably demonstrate this to others (Oakley, 1998, p.1241). However, implementing randomised controlled trials in real life settings is associated with a number of hazards. Oakley (1998) lists a few: low participation rates, high attrition, problems with informed consent, and unanticipated side effects of the intervention.

Despite the history of controlled experiments in education, the use of RCTs in educational research has declined. In the medical education literature over recent years, several educational researchers have stated that RCTs have important limitations in evaluating educational interventions (Prideaux, 2002, Norman and Schmidt, 2000). The first limitation highlighted was that of randomisation. While theoretically possible in educational research, Norman and Schmidt (2000) claim that sometimes it is not feasible or justifiable. They questioned
whether it is ethical to assign students to educational programs without offering any choice of learning method, adding that in research on educational interventions blind allocation is rarely possible. The second issue raised was that of control of variables. While it is theoretically possible to control for variables, in education the process depends on the context, and there are many variables. They also warned that efforts to control for complex variables may actually remove factors that are key in the success or failure of the intervention. The final issue of concern was the choice of appropriate outcome measures. The contemporary medical education literature has debated what outcomes of medical education indicate its efficacy. Improved patient health has been proposed as an appropriate measure of educational effectiveness (Harden et al., 1999, Prideaux, 2000). Unfortunately improved patient health is a complex measure, and subject to a number of variables beyond the control of the researcher. Prideaux (2002) suggested more theory-based research where outcomes predicted by educational theory could be tested with empirical work. He advised researchers to focus on significant outcomes that are amenable to intervention, so the findings could be attributed to the intervention, rather than be an artefact of the methods used. However, Torgerson (2002) was quick to respond, saying that as medical education is expensive, we have an obligation to rigorously determine the most effective educational methods. For Torgerson, the most effective method to evaluate the effectiveness of processes and practice in medical education was large, good quality RCTs. She dismissed claims that education is too complex and randomisation can’t control for the ‘myriad of
factors’ affecting outcome. In her words........it is precisely because of these influences, any of which can be associated with outcome, that randomisation is required in order to distribute them equally between the two treatment groups (Torgerson, 2002, p.1002). She was not concerned with the inability to double blind students and teachers in educational interventions, claiming that randomisation at baseline, not blinding, is what is needed for a robust trial. She dealt with the criticism that it is unethical not to give students choice of educational method in a trial, by claiming it was unethical to introduce a novel experimental educational method unless it has been shown to be at least as effective as traditional methods.

Power of a study is a crucial issue. Educational interventions, like health treatments, tend to have small effect sizes when they work. The average effect size of medical education has been reported as less than 0.5 (Albanese, 2000). To reliably detect an effect size of 0.5 (with an 80% power at a significance level of 5%) requires about 126 students in total, or 63 students/group (Armitage and Berry, 1994). It can prove difficult to achieve adequate power in a study despite efforts to gain ethical consent to participate from a cohort of students. Everyday impediments to student attendances, such as illness or outside employment, can negatively impact on student fulfilment of study requirements.
While self-report measures such as questionnaires have been criticised for recording what respondents say they do or believe at that moment in time, rather than what they actually do, they do offer an objective method of collecting information about people’s knowledge, beliefs, attitudes and behaviour (Boynton and Greenhalgh, 2004). Questionnaires may be administered to participants, or be self-administered. For the former, a person independent of the researcher should administer and collect the questionnaires to allow participants to respond freely without coercion. Self-administered questionnaires, accompanied by instructions and designed so respondents can complete them themselves, can be completed online, or distributed via email or a mail out to potential respondents. Postal surveys, as used in the IPE evaluation, have a number of advantages over administered questionnaires, in that they are able to reach high numbers of participants, and can increase the potential for anonymity and give respondents more time to respond. They also carry some disadvantages. Users of postal surveys run the risk of low response rates (Boynton, 2004), concerns about anonymity may arise when questionnaires are numbered to allow reminder letters to be sent to non-respondents, and if not completed immediately, surveys may ‘get lost’ in a participant’s pile of mail. If numbered questionnaires are used, respondents need reassurance that someone independent to the researcher has been appointed to handle all the data, including responses and reminders.
Boynton (2004) offers good advice on how to maximise the response rate: questionnaires should have a clear design and layout; participants should be invited to take part (preferably in advance); participants should be made to feel they are a stakeholder in the study; study aims and instructions should be clearly explained; participants should know how to contact the researcher with any questions about the study; a stamp addressed envelope should be included for a postal questionnaire; and the questionnaire should have a clear focus and purpose, and be concise. While low response rates can threaten reliability and validity, fewer questionnaires with good quality responses are preferable to higher numbers that are inaccurate or incomplete.

Questionnaire items may be closed or open-ended. Questionnaires with closed questions are........characterised by their use of pre-determined questions for all respondents (Shaughnessy and Zechmeister, 1990, p.79), and usually ask respondents to make a value judgement using a Likert-scale. Closed questions supply quantitative data for analysis using basic comparative statistics, and for each individual statistical test, an alpha level of 0.5 is traditionally used. They are traditionally associated with the key assumptions of the positivist paradigm. The data can be aggregated quickly, but the pre-determined and rigid structure of closed questions mean the researcher, not the respondents, sets the range of possible answers and thus the research agenda. As researchers may not consider all possible responses, this can lead to participant frustration (Boynton and Greenhalgh, 2004).
Questionnaires using closed questions, like other quantitative assessment or evaluation instruments, should be reliable and valid. As discussed in Section 1.6, reliable instruments yield consistent results from repeated samples and different researchers over time, and reliability is increased by standardisation of a measure. It is the differences between participants that should lead to differences in the results, rather than different understanding of the items by participants (Boynton and Greenhalgh, 2004). Valid questionnaires measure what they claim to measure. As in assessment, measures are taken to achieve validity but it is seldom achieved. Review of instrument items can give a rough estimate of content (or face) validity, and care should be taken with the wording of items to ensure they are unambiguous and the meaning is clear, and shared by all (ibid). Further support for validity (construct validity) can be obtained by triangulation, a process in which methodologies can be combined to give greater confidence to the findings (Stacy and Spencer, 2000). For example, ‘between methods triangulation’ is used when both closed and open questions are used in a questionnaire, and results from the quantitative research method are linked with those from the qualitative method respectively.

Open-ended questions, inviting free-text comments, can be added at the end or after individual items, allowing respondents to explain their answer, adding richness and perhaps, validity to the quantitative data (Murphy et al, 1998). This qualitative data can provide words that describe or explain phenomena that may be valuable for the evaluation process. In an inductive approach, the
researcher moves from data gathering to a general conclusions or principles. To facilitate this, the responses to the open-ended questions can be subjected to a content analysis to determine central themes. Again, differences between participants should lead to differences in the results, rather than different understanding of the items by participants, or in the analysis, different interpretations of the responses by raters (Boynton and Greenhalgh, 2004).

Content analysis

Content analysis is a research technique in which written and oral dialogue is objectively and systematically analysed, and described in a quantitative format. The quantitative information can then accurately represent the original qualitative information (Krippendorff, 1980). Thus it includes both quantitative and qualitative operations. This discussion will only focus on the type of content analysis used in the dissertation, conceptual analysis, used to establish the existence and frequency of concepts or themes. Themes that emerge from the text or transcript data are organised into mutually exclusive categories, defined to include all responses using the following method (Johnson and LaMontagne, 1993, Neuendorf, 2002): the transcribed data are read by the coder, to become familiar with the range of responses, and to note possible central themes; separate individual, complete thoughts within each response are identified and coded with initial categories; within responses, themes which refer
to the same idea are combined; coding of all responses is completed and from this the number of respondents making similar responses is calculated; and comparisons and tables are developed to express the themes and ideas of the participants.

Reliability is considered in terms of stability, reproducibility and accuracy. Stability refers to the tendency for coders to consistently re-code the same data in the same way over a period of time: reproducibility to the tendency for a group of coders to classify categories membership in the same way, and accuracy to the extent to which the classification corresponds to a standard or norm statistically (Gottschalk, 1995). Due to the human nature of researchers, coding errors can only be minimised, not eliminated. Usually efforts are made to improve the reproducibility of coding. According to Johnson and LaMontagne (1993), improved inter-rater reliability is obtained by the use of two independent raters, who each code at least 10-15% of the same responses.

Validity however, refers to how the categories correspond to the conclusions, and how the results can be generalised to a theory (Gottschalk, 1995). Validity of categories is achieved by using multiple classifiers to arrive at an agreed upon definition of a category. For example, for a category of ‘positive comments about the teaching method’, different students will express this in different ways. Coders must assign both explicit and implicit responses to this category. The most important challenge to validity is the conclusions inferred from the data.
The question is whether the conclusions follow from the data or are they explainable due to some other phenomenon (ibid). While large amounts of quantitative data can lead to reasonable conclusions, it often proves impossible to answer this question. As in assessment, steps are taken to improve validity, but it is seldom achieved. The generalisability of conclusions depends on both how concept categories are defined, and on how reliable the categories are (ibid). For a successful content analysis, it is critical to define categories which accurately measure the idea and/or items one is trying to measure, and that rules, which allow everyone to categorise and code the same data, the same way over a period of time (i.e. stability), are constructed. The conclusions and results are more sound if the specific categories, and methods used to establish all of them, are reproducible. Then, according to Gottschalk (1995) a study is said to be accurate.

Focus group methodology

Interviews have become popular way of collecting data, recognising that people don’t only respond to stimuli, but also act on the basis of their interpretations of the world around them and their experiences within it (Murphy et al, 1998). According to Murphy and colleagues (1992, p.112), the choice to use interviews to collect data can be simply interpreted as..........if you want to understand what people do, believe and think, ask them. Traditionally, interviews were one-
to-one between the researcher and one respondent, but increasingly, group interviews or focus groups are being used. Focus group interviews or discussions have become a popular methodology in medical education research as they can be used to understand a group of people’s beliefs, opinions and attitudes about a topic of interest (Vermeire et al, 2002). They are one type of group research defined by the following statements........a group is a number of interacting individuals having a community of interest, the term interview implies........the presence of a moderator who uses the group as a device for eliciting information, and the term focus in the title implies........the interview is limited to a small number of issues (Stewart and Shamdasani, 1990, p.10).

Contemporary focus groups usually involve a relatively homogeneous group of 6-12 people, whose responses to prepared questions are probed and induced by a trained moderator in a relatively short time (Vaughn et al, 1996). The purpose is to gather a range of opinions about issues, rather than build consensus amongst people (ibid). Focus groups have several advantages compared to individual interviews. These include synergism (the group interaction results in a wider bank of data), snowballing (the statements of one respondent can initiate a chain reaction of additional comments), stimulation (group discussion generates excitement about a topic), security (comfort and candid responses are encouraged by the group) and spontaneity (responses are more spontaneous and genuine as participants don’t have to answer every question (Stewart et al, 1990, Vaughan et al, 1996).
Focus group interviews assume that people are valuable sources of information about themselves. There is a greater emphasis on the participants’ points of view compared to interviews as the participant-inquirer interaction is replaced by the interaction between participants (Vaughan et al, 1996). Some of the many advantages have been described above. Important limitations include the fact that individual responses are not independent of one another, limiting generalisability of results; individual behaviour may be subject to group influence; participants may modify their opinions after interactions; and dominant or opinionated group members may bias the results (Stewart et al, 1990).

The method is compatible with the three key assumptions of the constructivist paradigm (Vaughn et al, 1996). As explained in Section 2.3, a fundamental tenet of this paradigm is that multiple views of reality can exist. Individuals devise socially constructed realities as they attempt to make sense of their experiences, which can be and usually are shared. This does not make them more real but simply more commonly assented to (Lincoln and Guba, 1985). It also recognises that findings arise from the interaction between the inquirer and respondent, and this relationship can potentially add depth and dimension to the knowledge gained (Vaughan et al, 1996). However, care is needed with the inquirer-respondent interaction, as the moderator may introduce bias into the study, by knowingly or unknowingly providing cues to participants about what types of responses and answers are desirable (Stewart et al, 1990). Finally in
this qualitative tradition, there is an assumption about the nature of truth. The latter is explained by describing a particular set of concepts, in relation to a particular context (which must be documented). The quality of focus group work can be critically appraised by the criteria already described above for qualitative research. Trustworthiness can be improved by providing excerpts from the interview for each discussion topic, and providing examples in sufficient detail; and sampling should include documentation of the recruitment, how decisions were made about the number of groups planned and performed, the characteristics of, and the relationship between, the participants; and the context and setting (Vermeire et al, 2002). For some, ‘respondent validation’ or ‘member checking’ is the ultimate check on credibility of a study’s findings (Lincoln and Guba, 1985), but others see this is more of an opportunity for ‘reflexive elaboration’ of the original analysis (Emerson, 1981). The fundamental problem of interviewing, whether individual or in groups, is that the context of report collection places constraints on interviewees’ reports (Murphy et al, 1998).
2.5. Ethics

Most major research and government institutions, such as the National Health and Medical Research Council (NHMRC), have agreed the following set of guidelines for ethical practice, to regulate research involving human subjects (as cited in Murphy et al, 1998, p.148):

- Informed consent i.e. the subjects of the research fully understand what participation in the study will involve, and that they have given their consent to participate without coercion
- Confidentiality i.e. all information given by, or obtained about the research subjects, should be kept in confidence
- Subjects rights and welfare i.e. subjects should be aware that investigators may, quite unintentionally, introduce unnecessary or unacceptable hazards, or fail to provide adequate safeguards
- Subject risk-potential benefits ratio i.e. unknown or foreseeable risks to subjects must be outweighed by the probable benefits that may be experienced by them and/or humanity by their participation in the project

These guidelines are derived from four ethical principles identified by Beauchamp et al (1982) as relevant to ethical practice in research involving people. Autonomy or self-determination (we should respect the values and decisions of other people) is the principle that underlies informed consent. The principles of non-maleficence (it is wrong to intentionally inflict harm on
another person) and beneficence (we have an obligation to remove existing harms and to confer benefits on others) inform the practice of risk-benefit analysis. The guidelines are also informed by the principle of justice (people who are equal in all relevant respects should be treated equally). While these guidelines are mostly derived in the context of biomedical research, the ethical principles of autonomy, justice and protection of the participant from harm apply in all kinds of research (Murphy et al, 1998).

Random allocation to control or interventional groups raises the same ethical issues in both biomedical and educational research, but in the latter there tends to be less risk of serious or irremediable harm compared to biomedical research. However in educational research, participants may be at risk of harm to their psychological and emotional wellbeing, or their professional or legal position; and instead of risk occurring during the research, the greatest risk may occur when the results of the study are disseminated (ibid). As the outcomes of qualitative research are potentially more complex or unpredictable, some particular ethical problems may arise. For example, it may not be possible for consent to be fully informed at the beginning of a study, and researchers must reflect on all possible ethical implications of the proposed work, before the study, and whether it is indeed ethical to proceed. Care should be take to minimise any residual risks to research participants (for example, procedures to anonymise the data), so the potential for psychological or emotional harm or negative consequences of publication of the study are reduced.
References

Albanese, M. (2000), “Problem-based learning: why curricula are likely to show little effect on knowledge and clinical skills”, Medical Education, vol. 34, no.9, pp.729-738


Bogdan, R. and Bicklen, S. (1982), Qualitative research for education: an introduction to theory and methods, Allyn and Bacon, Boston


Johnson, L.J. and LaMontagne, M.J. (1993), “Research methods: Using content analysis to examine the verbal or written communication of stakeholders within early intervention”, *Journal of Early Intervention*, vol. 17, no.1, pp.73-79


*Applied Social Methods Series*, vol. 20, Sage Publications, Newbury Park, CA


3.1. Statement of authorship

Linking cardiovascular theory to practice in an undergraduate medical curriculum

*Advances in Physiology Education* (2001), vol. 25, no. 4, pp.193-201

**Hudson, J.N.** (Candidate)

Designed, developed and implemented the Case Based Teaching initiative, and the evaluation. She also wrote the manuscript and acted as corresponding author

Signed……………………………………………………Date………………..

**Buckley, P.**

Conceived the idea of the Case Based Teaching Tutorials, and contributed to the thinking and critique of the evaluation and manuscript

Signed……………………………………………………Date………………..

**McMillen, I.C.**

Conceived the idea of the Case Based Teaching Tutorials, and contributed to the thinking and critique of the evaluation and manuscript

Signed……………………………………………………Date………………..
3.2. Linking cardiovascular theory to practice in an undergraduate medical curriculum

Hudson, J.N.¹, Buckley, P.² and McMillen, I.C.²

¹The Medical Education Unit, School of Health Sciences
²The Discipline of Physiology, School of Molecular and Biomedical Sciences

The University of Adelaide, South Australia 5005, Australia

Advances in Physiology Education (2001), vol. 25, no. 4, pp.193-201
Hudson, J. N., Buckley, P. and McMillen, I.C.
Linking cardiovascular theory to practice in an undergraduate medical curriculum
Advances in Physiology Education 2001, v. 25 no. 4 pp. 193-201

NOTE: This publication is included in the print copy of the thesis. It is also available online to authorised users at:

http://advan.physiology.org/cgi/content/abstract/25/4/193
Medical education: the challenge of linking theory to practice       J.N. Hudson

4.1. Statement of authorship

An evaluation of Case Based Teaching: evidence for continuing benefit and realization of aims

Advances in Physiology Education (2004), vol. 28, no. 1, pp.15-22

Hudson, J.N.  (Candidate)  
Contributed to the conception and design of the evaluation, interpreted and analysed the data, and implemented the evaluation. She also wrote the manuscript and acted as corresponding author
Signed……………………………………………………Date………………..

Buckley, P.
Contributed to the conception and design of the evaluation, helped in data interpretation and manuscript evaluation
Signed……………………………………………………Date………………..
Chapter 4

4.2. An evaluation of Case Based Teaching: evidence for continuing benefit and realization of aims

Hudson, J.N.¹, Buckley, P.²

¹The Medical Education Unit, School of Health Sciences
²The Discipline of Physiology, School of Molecular and Biomedical Sciences

The University of Adelaide, South Australia 5005, Australia

Advances in Physiology Education (2004), vol. 28, no. 1, pp.15-22
Hudson, J. N. and Buckley, P.
An evaluation of case based teaching: evidence for continuing benefit and realization of aims
Advances in Physiology Education 2004, v. 28 no. 1 pp. 15-22

NOTE: This publication is included in the print copy of the thesis. It is also available online to authorised users at:

http://advan.physiology.org/cgi/content/abstract/28/1/15
5.1. Statement of authorship

Computer-aided learning in the real world of medical education: does the quality of the interaction with the computer affect student learning?

*Medical Education* (2004), vol. 38, no. 8, pp.887-895

Hudson, J.N. (Candidate)

As solo author, designed and implemented the evaluation, analysed and interpreted the data, wrote the manuscript and acted as corresponding author.

Signed……………………………………………………Date……………………

Medical education: the challenge of linking theory to practice J.N. Hudson
5.2. Computer-aided learning in the real world of medical education: does the quality of the interaction with the computer affect student learning?

Hudson, J.N.

The Medical Education Unit, School of Health Sciences

The University of Adelaide, South Australia 5005, Australia

*Medical Education* (2004), vol. 38, no. 8, pp.887-895
Hudson, J. N.
Computer-aided learning in the real world of medical education: does the quality
of the interaction with the computer affect student learning.
Medical Education, 2004, v. 38 no. 8 pp. 887-895

NOTE: This publication is included in the print copy of the thesis.
It is also available online to authorised users at:

http://dx.doi.org/10.1111/j.1365-2929.2004.01892.x
6.1. Statement of authorship

Evaluating the impact of moving from discipline-based to integrated assessment

Medical Education (2004), vol. 38, no. 8, pp.832-843

Hudson, J.N. (Candidate)
Conceived the idea of the integrated practical examination and its evaluation, organized the 1999 examination, and co-ordinated the project from its inception to completion. She also wrote the manuscript and acted as corresponding author
Signed……………………………………………………Date………………..

Tonkin, A.L.
A.L. Tonkin organized the 2000 examination and analysed the data for both years. She also made a contribution to writing of the paper.
Signed……………………………………………………Date………………..
Chapter 6

6.2. Evaluating the impact of moving from discipline-based to integrated assessment

Hudson, J.N. and Tonkin, A.L.
The Medical Education Unit, School of Health Sciences
The University of Adelaide, South Australia 5005, Australia

Medical Education (2004), vol. 38, no. 8, pp.832-843
Hudson, J. N. and Tonkin, A. L.
Evaluating the impact of moving from discipline-based to integrated assessment.
Medical Education, 2004, v. 38 no. 8 pp. 832-843

NOTE: This publication is included in the print copy of the thesis.
It is also available online to authorised users at:

http://dx.doi.org/10.1111/j.1365-2929.2004.01893.x
Chapter 7

Discussion

7.1. Introduction

This chapter analyses the outcomes of the published papers, which comprised Chapters 3-6, drawing conclusions in relation to the background literature, and contemporary context of medical education. Section 7.1 outlines the structure of the chapter. After briefly summarising the contemporary setting of medical education, section 7.2 presents the key findings of the published papers, including their strengths and weaknesses and contribution to new knowledge. Section 7.3 discusses the limitations of the evaluative methods used in the studies and section 7.4 suggests future research directions. Finally, section 7.5 presents the references discussed in the chapter.

7.2. Summary, key findings and their contribution to new knowledge

To remain useful, medical education must be responsive to changing values and expectations. As described in Chapter 1, medical education has undergone
various transitions but changes instituted in 1920 were significant, generating the theme of this dissertation. Medical education has continued to change, especially over the past 30 years. Notably, information technology has offered exciting new virtual environments in which to deliver professional education, and conceptions of professional competence have changed. A contemporary definition of an individual’s level of competence in an area of practice is an integrated one, which includes knowledge, skill and acumen (Kane, 1992). This has had implications for learning environments, as well as assessment. Recognition that assessment is a powerful drive to learning meant that the same values should guide the design of both learning and assessment activities.

Internationally, there are schools still expressing the ‘Flexner’ model of medical education, some have adopted the SPICES model and some are testing a new model put forward by Bligh and colleagues in 2001. The driving force of the latter approach is clinical education, in response to a perceived need of greater clinical experience among medical students. They recommend a product-focused curriculum and practice-based education where possible, and the development of professional behaviours and clinical experience underpinned by the underlying basic and clinical sciences (Bligh et al, 2001). A feature of this, and most contemporary models, is problem-based learning, introduced to integrate the basic, human and clinical sciences in the context of a virtual case. However, PBL has also been accused of maintaining the pre-clinical/clinical divide, especially when problem-based learning is stressed in the early years and problem-based
solving and management, in the latter (Margenson, 1999). The early introduction of clinical skills teaching has made some contribution to a seamless curriculum in clinical skills, and improved student satisfaction with undergraduate studies (Bradley and Bligh, 1999). Despite these changes, recent literature has suggested that all is not right. For example, fourth year medical students who demonstrated competency in the directed examination of organ systems, were unable to correctly apply these examination skills to work up clinical cases (Wilkerson and Lee, 2003). Teaching clinical skills fosters ‘acting like a doctor’, and PBL fosters ‘thinking like a doctor’ but as Schön (1987) demonstrated after thirty years of careful study, doctors think and act at the same time. Despite a considerable body of evidence in professional education, from eminent researchers such as Eraut (1985), Schön (1987), and Bruner (1996), integration of theory and practice had not been achieved in many medical curricula by the latter part of the twentieth century.

In the mid- to late 1990s at the University of Adelaide, the undergraduate medical curriculum was a version of the SPICES model, a hybrid of both traditional and problem-based learning strategies. Science was valued, as described above (Bligh, 2001) for ‘application, interpretation, imagination and integration’. The latter was a key issue for the sciences that underpinned the practice of medicine, namely human physiology and anatomical science. Learning in human structure and function was co-ordinated rather than integrated, and assessment tended to be discipline-based. Initiatives were needed, not only to integrate structure and
function, but also to integrate the basic sciences with contemporary clinical practice in learning and assessment settings. In addition, there were calls for more self-directed learning, and to embrace and evaluate the potential of electronic learning. The published initiatives were the response to these curriculum directives. Their strengths and weaknesses, and outcomes, will be summarised and discussed in turn.

_The CBT innovation, student and tutor perceptions of its benefits and success in realizing aims_

The strength of CBT lay in its foundation in Schön’s (1983, 1987) model of reflective practice. It revealed how science interacts with other factors in the context of medical decision-making. Students learned by thinking and acting like doctors at the same time. CBT simulated rather than reflected the actual practice world, as peers rather than patients, were the objects of student practice. It was free of the pressures and distractions of a busy hospital ward, and there was no risk of harm to the infirm. Students solved real-world clinical problems, using a range of resources, described by Bines and Watson (1992) as the tools, language and appreciation of their ‘trade’. Professional knowing was seen in terms of ‘thinking like a doctor’, and students learned how to reason their way through their general knowledge to apply to a particular case, as competent practitioners do for a case that is problematic, when familiar categories and ways of thinking fail (Bines and Watson, 1992). By using reflection-in-action to reach
a diagnosis, they started to acquire rules and methods of their own for future use. Learning, rather than teaching, was key as it prepared students for situations of uncertainty about the answer to a problem. It was part of a continuum towards the real world practicum of clinical education (in the hospital or community setting), where an experienced clinician would ‘continue reflective conversations’ to facilitate the acquisition of this important part of expertise or competence (Bines and Watson, 1992, Schön, 1987).

Another key to the success of the CBT practicum was the tutor, who by advising, questioning and critiquing, helped students through the transition from lay-person to medical practitioner. Novice students were able to observe successful performance of a new task prior to attempting it themselves, as well as receive positive encouragement to achieve realistic tasks. As this was situated early in the learning continuum, it fostered a positive perception of self-efficacy (Bandura, 1986). Tutors believed that it was the non-threatening environment that helped students gain confidence for their first ‘real patient’ encounters. In the comfort of the small group setting, students were also important to each other and sometimes played the tutor’s facilitatory role. In small groups, students were immersed in a simulated practice world, learning by exposure and immersion (Bines and Watson, 1992). According to Schön (1987), this background learning often proceeds without conscious awareness, although a student may become aware of it later on, when moving into a different setting.
This point was also raised by one of the tutors in the longitudinal evaluation of CBT, described in Chapter 4.

Learning by exposure and immersion takes time and therein, lays one of the weaknesses of the CBT initiative. In the longitudinal evaluation, students requested more time for CBT, and notably most requests came from the year 2 students. Greater time allocation would have allowed individual students to practice and receive more feedback on their thought processes and skills, important when CBT was valued by students as their first introduction to clinical skills. The greatest support for CBT came from third year medical students, at the crossroads between their preclinical and clinical environments. Fourth year students, now encountering real patients, had moved on to developing their skills in the hospital environment. Traditionally, it was the hospital environment where novice students gradually developed into experts by acquiring skills as they gathered experience. Theory was not separated from practice in this ‘apprenticeship’ environment. However, it had problems (already discussed) and was less available with a rise in student numbers paralleling a decline in hospital patient numbers. CBT provided a solution, building towards a seamless curriculum.

The longitudinal evaluation of CBT (Chapter 4) was conducted when CBT was a stable component of the ongoing medical curriculum. While the smaller formative evaluation (Chapter 3) also investigated whether educational goals had
been met (as defined by the first definition in section 2.2.), it took place during the CBT development phase, to gauge what improvements could be made. In the CBT evaluations, both quantitative and qualitative methods were used, the former to quantitate the perceptions of the whole cohort, and the latter to determine why or how individuals held their views. In the formative evaluation after the initial introduction of CBT (Chapter 3), over ninety per cent of second year medical students believed that this initiative had been a valuable learning experience relevant to its stated aims, helping them understand how and why clinicians interview and examine patients. A selection of student answers to the open-ended question, ‘what aspects of the tutorials did you like best’, is reiterated below:

- The practical helps put other aspects of our learning in place
- Gave us a chance to try out physical examination and understand why each was done
- The practical aspect was very interesting and made what we were doing in the lectures seem more applicable to the requirements of my profession
- The practical aspects of human structure and function topics were covered, which increased our understanding and recall of the theory
- The hands-on work showed the principles and theory behind it

In the subsequent longitudinal evaluation (Chapter 4), second, third, and fourth year undergraduates and their tutors believed that CBT met its three major aims. CBT was valued as a learning experience that not only integrated physiology with related basic sciences and clinical medicine, but importantly linked
students’ developing knowledge of theory to practice. Their spontaneously offered comments reflected the same values espoused in the initial evaluation, and illustrated in the quotes above. This supported the validity of both evaluations (between-data-triangulation for convergent validity). The qualitative data derived from the open-ended questions on each questionnaire added to the validity of assumptions made from the quantitative results (between-methods-triangulation for convergent validity).

The Computer Aided Learning and the most effective format for learning gains, in terms of ability to retain and apply knowledge

As discussed earlier, an understanding of human structure and function underpins problem solving in clinical practice. For most medical practitioners, human structure will be viewed as surface or radiological anatomy. Surface anatomy can be well integrated with practice during the acquisition of physical examination skills, but other learning activities are needed to bridge the theory-practice gap, to foster expertise in the latter. As computer presentation is well suited to visually intensive topics such as radiology, CAL was an ideal learning environment to link students’ developing knowledge of gross and cross-sectional anatomy with its major use in clinical practice.

Novel strategies such as computer-aided learning are expensive to design and implement so evidence of their effectiveness is needed to inform decisions about
their use in medical education. Most studies have compared CAL to more traditional learning formats rather than determine the most effective CAL format to use. Although the ultimate aim of medical education may be improved patient health, it is correct to say that no direct association can be easily inferred between a CAL intervention and patient health. However, the CAL study focused on a significant outcome, student learning, so the findings could be attributed to the intervention. This, together with the fact that the outcomes were predicted by educational theory and tested with empirical work, was the strength of the study. To maintain a ‘laboratory of life’, with greater external validity (Smith, 2002), nothing more was done than the randomised allocation of students at baseline, to experimental and control groups.

The mainstream definition of evaluation (as described in Section 2.2) fits the model used to evaluate the CAL initiative. Rather than prove a particular point of view, it aimed to improve the use of e-learning by demonstrating its most effective format. As subjects were randomly allocated to different groups at baseline, the interaction between the researcher and the subjects was minimised, meeting the criterion of neutrality. For a research question that asked ‘how much’ better is one learning format than another, a randomised controlled trial was appropriate. Similarly, the intention-to-treat analysis that was used to make the judgement was appropriate, as the research question referred to whether any format does improve learning, rather than whether it can. The inability to double blind students and teachers in the intervention was not considered to be a
problem as students were randomised at baseline, the latter being key for a robust trial (Torgenson, 2002). Moreover, the assessor, who marked all the tests of student learning before and after the CAL intervention, was blind to group allocation. The fact that students couldn’t choose to which educational method they were allocated was not unethical. Rather it was seen to be unethical to introduce a novel educational method unless it was at least as effective as traditional ones (ibid).

The CAL study in Chapter 5 resulted in some useful outcomes. It showed that medical student users of an interactive CAL tutorial had significant learning gains compared to non-CAL users, but these gains were not superior to those achieved from non-interactive CAL. It also demonstrated the difficulty of conducting randomised controlled trials in the real world of medical education, and the importance of using a valid assessment tool when evaluating learning outcomes from educational interventions such as CAL packages.

The major weakness of the study was its power. Due to cohort size, and limitations placed on student participation such as competition with students’ professional and personal workloads, adequate study power was not achieved, according to the criteria described in Section 2.3. (Armitage and Berry, 1994, Albanese, 2000).
Another strength of the study was the assessment used to measure learning. It tested authentic professional tasks, and the ability to apply and retain knowledge rather than just recall. Different response formats were used in the test to ensure that any possible correlation between assigned method of learning and success in an assessment format, didn’t introduce bias into the study. Reliability was also improved by the use of one assessor to judge all student responses to assessment items. While greater validity may have resulted from consensus between several judges of student answers, cost limitations excluded this. However, validity gains were made by critical review of both questions and ‘ideal answers’, by educational and content experts, as well as blueprinting against tutorial learning objectives. Validity gains were also made in terms of expectations of the test, as experts scored higher than third year medical students, who in turn scored higher than complete novices. Finally, the test was useful as a source of formative feedback on learning and it showed students that practice underpinned by theory, and understanding and retention of knowledge, were valued.

*The development and implementation of a novel assessment, The Integrated Assessment (IPE), and its utility as an assessment instrument, as judged by students and staff*

Using the variables in van der Vleuten’s (1996) model, the final study in this dissertation revealed that the IPE was a useful test at a time of curriculum and assessment reform at the University of Adelaide. A combination of quantitative
and qualitative methods provided rich data for continuous quality improvement. The qualitative data helped to elaborate on the problems identified by the quantitative data, for example in relation to student reluctance to abandon the ‘rote learn and dump’ style of learning for the assessment. Conversely, while the qualitative data revealed individual strong criticism or praise for the intervention, the quantitative data revealed whether this was representative of the whole cohort’s opinion or not.

The WHO definition of evaluation from section 2.2 captured well the evaluation of the integrated practical assessment, one of its strengths being its comprehensive nature. The evaluation had several major positive outcomes. Firstly, the large investment of time and effort required to reform the assessment practice was rewarded by gains in assessment utility. Good reliability was achieved, by gathering a sufficient sample of performance and scoring it with some level of standardisation and structure (van der Vleuten, 1996). The IPE, an innovative version of the OSCE, tested an integrated view of competence, reflecting the richness of professional practice. Acknowledged improvements in validity were achieved by linking theory to practice and subjecting test items to an extensive review process. The gains in content validity were made by using experts and consensus, during the development and review process, and mapping of the assessment items against expectations of a third year undergraduate medical student (in the absence of specified Learning Outcomes for the course).
Agreement on answers, as well as questions, by a wide range of professionals contributed to this improvement.

The IPE also had outcomes in terms of consequential validity, i.e. the effect of assessment on learning, and how the results are used (Hutchinson et al, 2002). Snyder (1971), amongst others, has shown that assessment is a powerful drive to learning. In previous third year examinations at Adelaide University, some discipline based examinations rewarded students for rote learning and short-term recall of information, and the large number of exams were often structured in such a way, that they were in competition with each other. Although the reform of assessment practice wasn’t universally successful in changing student learning behaviours, this should be more achievable in the new integrated curriculum that followed. The study did however promote greater understanding, and improved performance, of assessment practice for staff. It also showed that when moving from a discipline-based to an integrated curriculum, assessment could be used to drive integration of learning and teaching; and when efforts are made to gain the support of both staff and students prior to the change, the process is facilitated. Acceptability by staff and students was a critical issue. For the assessment change to survive, we had to address staff concerns, provide them with support and information, and ask those who participated in the focus group to endorse conclusions reached from the data, improving their credibility. This also gave participants a further chance to withdraw consent prior to dissemination. In the interests of reflexivity, a
facilitator with no vested interest in the study was chosen to moderate the focus group, rather than the researchers. However, the moderator was known to the participants to exploit the benefits of well-facilitated interaction between the members of a group.

All the activities described in Chapter 6 are resource intensive, and thus costly. While not extensively researched, the implementation costs may be justified, in terms of the perceived benefits for learning outcomes. The study would have been strengthened by greater attention to this issue. The results of the evaluation were subsequently used to improve the relevance and integration of the third year curriculum, and the assessment program in the new curriculum at the University of Adelaide includes a structured integrated practical examination during the first three formative years of the course.

7.3. Methodological limitations

Each of the studies in this thesis followed good research practice, as defined by Murphy et al (1998): the design and methodology were chosen according to the research topic or question; sampling was done systematically, in some instances tempered by pragmatics; the method of information gathering and analysis was described; the results addressed the research question and trustworthy conclusions made; and the meaning of the findings were addressed and interpreted in the light of existing knowledge. There are however some
methodological weaknesses that must be considered when interpreting the results.

Firstly there are some issues in relation to the questionnaires. Those used in both the CBT and IPE studies may have benefited from the use of a qualitative approach first. As Boynton and Greenhalgh (2004) advise, this can help to explore the territory and map key areas for further study. A focus group discussion with the students (and tutors) could have flagged up potential issues and help predict the range of possible responses, leading to better questionnaire design and useful results. Time and cost led to the compromise used. The qualitative data (open-ended questions) did however complement the quantitative data (closed questions), capturing a more complete portrayal of the data under study (Lev, 1995), adding greater confidence to the findings (Stacey and Spencer, 2000) in both the CBT and IPE evaluations. In retrospect, ‘between-methods-triangulation’ could have been used more in the interest of completeness (Murphy et al, 1998), to draw attention to the different perspectives of individual students. For example, in the formative CBT evaluation (Chapter 3) unique responses of students are illustrated in Table 1. Only a few students requested ‘expert tutors’, and one student requested that the CBT tutor do ‘less didactic teaching’. It would have been useful to explore the meaning of these comments, and how they may have influenced CBT effectiveness in meeting its aims. This could then be accounted for, not only in
improving the learning environment, but also in its subsequent longitudinal evaluation.

In all questionnaires questions were solely stated in the positive. It was a format familiar to students at the University of Adelaide in evaluations conducted by the Advisory Centre for University Education (ACUE), and students had been noted to freely express both positive and negative opinions despite this format. The ACUE was used to collect and analyse some of the initial data in the interests of ‘neutrality’, however adopting their preferred style may have introduced a response bias.

The questionnaires were administered by an independent administrator or sent by post to reduce the influence of the researcher. While postal surveys do run the risk of low response rates, in the IPE evaluation it was used for pragmatic reasons (some of the students had moved to various clinical settings). For each study, decisions on sampling timing and mode were made on the basis of when the majority of students (or tutors) could be accessed. Despite efforts to improve the response rate (as detailed by Boynton, 2004), it was less than adequate for the IPE postal survey. There is a risk that the opinions expressed by respondents did not represent those of the whole cohort. This was the greatest challenge to its validity, as conclusions were reached from the data of only those who responded.
Poor response rates also challenge the reliability of a measure, as different results may be obtained from the same sample, but with a different response rate, on another administration. We had to accept that in the context and timing of the studies, maximum effort had been exerted to minimise possible error and/or bias due to this, and that good quality responses from students who wished to express their views were preferable to high numbers of incomplete or inaccurate responses. It should be added however that in the CBT longitudinal evaluation, the qualitative data revealed that some second year students may not have understood the item in relation to aim 3 (21 percent of respondents were independently judged to have misunderstood this question). Inconsistency in how respondents understand an item can also challenge the reliability of a questionnaire. The between-methods-triangulation used didn’t improve this weakness as the quantitative question related to the same expression of the aim. Thus less reliable conclusions can be drawn from the sample of second year students in relation to aim 3 (Chapter 4). However reliability (in terms of the reproducibility) of the qualitative data from each questionnaire was improved by using more than one coder or rater to interpret the responses, in the content analysis. The use of at least two independent raters, who each coded all the responses (exceeding the ten to fifteen per cent advised above), contributed to inter-rater reliability.

While Cronbach’s alpha is the most common method of measuring internal consistency, in the IPE evaluation it may have overestimated the reliability of
the assessment. Use of a generalisability coefficient (g), rather than Chronbach’s alpha, would likely have been a more accurate measure of reliability, as it accounts for all the sources of variability (examiner, patient and rating method) that were influential simultaneously in the integrated exam (Swanson et al, 1987).

All staff contributors to the assessment reform were consulted on the acceptability of the IPE. Although two key members were unavailable on the day agreed by consensus, opinions of those sceptical, and those more accepting, of the change were still obtained. The data from the focus group interview was highly contextual and as previously explained, was not generalisable. Excerpts of the staff responses to the interview questions would have improved the trustworthiness of the conclusions drawn from the data, as was done with some of the student responses to the open-ended questions in the CBT evaluation (Chapter 3). However, the confines of publishing the manuscript meant that all aspects of the comprehensive evaluation couldn’t be presented in Chapter 6.

The discussion in Chapter 5 describes and critiques in some detail, the measures taken to address the reliability, validity and educational impact of the test used to assess the CAL, as well as the internal and external validity of the evaluation. As discussed in Section 7.3, the major limitation of the CAL evaluation was the power of the study. It proved difficult to recruit a sufficient number of students from a one-year cohort of medical undergraduates at the University of Adelaide
given the limit on class size, as well as other factors such as ethics, and curriculum overload. Recruitment of a larger number of subjects from different years, or different institutions may have added to study power, but as each cohort has different levels and types of experience, the internal validity of such a study would suffer. Possible strategies to address this are presented in section 7.4.

Some qualitative data would have added to the findings from the CAL evaluation. Questions such as: ‘why did students spend more time on the didactic, and less on the free-text response, tutorial; or ‘why do students prefer one CAL format over the other' would have potentially provided answers, instead of assumptions about the responses to these questions. Addressing this, as well as other potential barriers to participation, could have led to a study of greater power, where conclusions about the effectiveness of CAL tutorial formats would have been more relevant and valid.

For each of the studies, the researchers were guided by the ethical guidelines summarised in Section 2.5, with ethical approval being granted by the internal departmental ethics committee. The issue of potential implications of study publication was relevant in relation to the focus group findings. Care was taken to anonymise the data so comments couldn’t be attributed to an individual group member. However it was a small group, the members working on the
assessment reform were known at the institution, and some may have felt professionally challenged by the collective stance being published.

As the statistical methods used in analysing the evaluation data from all the published studies have been subject to international peer review and accepted as appropriate, they will not be critiqued in this discussion.

7.4. Future directions

Over the past one hundred years, medical education has focused on teaching diagnosis and management but in 2004 society is asking more of doctors. Medical schools are being required to focus on training doctors who will practice patient-centred medicine, and see health as an interplay between medical expertise and the patient in the context of his/her illness (Jamshidi and Cook, 2003). Diagnosis and management however, remain important, but the emphasis has shifted from teaching to learning (Schuwirth, 2003). This thesis has made a contribution to professional education by determining the outcomes of learning and assessment initiatives that were based on contemporary educational theory. Some of the studies would benefit from further research.

Assessment should continue to be seen as a means rather than an end to learning. For the IPE, validity gains could be achieved by getting students at different stages and medical practitioners to do the test, with the expectation that experts
will score higher than non-experts in a valid test. Construct validity would be addressed by a study correlating student scores on the IPE, with those from the other integrated tests in the third year assessment program, the written MCQ and MEQ exams; and predictive validity by comparing the IPE scores of individual students with their performance as they move on through their training. However, rather than focus on just one exam, further studies should address the utility of the whole assessment program, how scores are combined and how the standards are set. Patients, increasingly asking to be included in the societal conversation about health and health care (Cribb, 2000), should be players in student assessment, and program evaluation.

According to Parker (as cited in Noonan, 2002, p. 59), the computer........is the black bag of the future, if used the right way. Technology is transforming the practice of medicine and patient experiences, as well as the way tomorrow’s doctors are being trained. Uses range from realistic medical simulation (Gordon et al, 2004) to reusable learning tutorials. Just like face-to-face learning environments, CAL needs to be evaluated. The utility of interactive computer-learning environments still needs to be tested, using a study of greater power. This could be achieved with a study at an institution with larger student numbers, or by conducting a series of smaller studies in the real world of medical education at a number of different institutions. Further work could extend to computer-based simulation to help students identify the essential aspects of the physical examination in specific clinical cases, one solution to the previously reported
problem of student difficulty with clinical decision-making (Wilkerson and Lee, 2003). An interactive CAL tutorial in the free-text style would allow immediate feedback on the aspects of the history and physical examination individual students chose as essential.

To continue the theme of linking theory to practice, it would be interesting to further develop the ‘practicum’ and explore how it may foster the development of competence in a more seamless curriculum. There is potential for using high-fidelity patient simulators, which Gordon et al (2004, p. 23) described as a "natural framework for the integration of basic and clinical science in a safe environment." They used simulation to bring ‘life’ to their PBL case material, as well as using it in ‘transition’, and the clinical years. However, this doesn’t de-emphasise the need for the human touch.

Future directions, either in a natural or simulated setting, include a study investigating the decision making of medical students and doctors (practitioners) at different stages of their training, to determine how their understanding of the meanings of the history, examination and investigation findings are used in professional action. According to Schön (1987), practitioners bring their knowledge and experience to every new situation. If the results from the physical examination don’t match the symptoms, or if the latter don’t match the investigation findings, practitioners are surprised. This leads to reflection-in-action and further action. Apart from potentially leading to a diagnosis, reflection
on the new actions also expands the knowledge and experience that the practitioner brings to the next professional action or encounter. Based on this model of reflective practice, the study would focus on specific decision making episodes that went well, asking participants to reflect on the decisions they took, and why they took these decisions, prompting them to consider what theory (basic and human sciences, as well as ‘craft’ knowledge) they used, what features of the context (attitudes, expectations, policies) influenced their action, and what personal or societal values played a dominant role. The participants’ background and style of medical education would allow comparison of different curricula. For students who graduated from the University of Adelaide over the past few years, the CBT learning experience may form part of their reflections that inform future decision-making.
7.5. References

Albanese, M. (2000), “Problem-based learning: why curricula are likely to show little effect on knowledge and clinical skills”, Medical Education, vol. 34, no. 9, pp.729-738


Schuwirth, L. (2002), “Can clinical reasoning be taught or can it only be learned?”, *Medical Education*, vol. 36, no. 8, pp.695-696


http://bmj.bmjjournals.com/cgi/eletters/324/7330/126


