University of Glasgow
Faculty of Medicine
School of Nursing and Midwifery and Department of Public Health.

The Instructional Design and Evaluation of a Multimedia Program to Help Mentors Develop Skills in Assessing Student Nurses' Clinical Performance.

Charles Docherty, MN (Glasgow), RGN, RNT, DipN (London)

Thesis submitted to the Faculty of Medicine, University of Glasgow, for the degree of Doctor of Philosophy (Ph.D.)

May 2002

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

Background

The major organisational change accompanying the introduction of Project 2000 and the incorporation of nurse education into higher education exacerbated perennial concerns over the validity and reliability of student nurses' clinical assessments. Problems identified in the literature review were lack of consistency and poor reliability of ratings, instructor subjectivity, questionable validity of assessment tools, poorly utilised assessment criteria and reluctance of assessors to negatively rate or fail students. These were partly attributed to difficulties inherent in clinical assessment, and partly to the inadequate preparation of assessors. Further analysis of the assessment of clinical practice identified contradictory and confusing terminology, such as competence and performance. The process of assessing practice was found to be devoid of an agreed theoretical basis.

The historical origins of mentorship and its introduction to nurse education with Project 2000 were explored, with evidence from the literature suggesting that the attitudes, knowledge, anxiety, and self-efficacy of mentors can influence how well their role is performed, in particular, their assessment of students. Meanwhile, a revolution in learning technology offered multimedia as a potential solution to the logistical difficulties of providing a quality preparation for mentors. Evaluative studies suggested that computers had the potential to improve the effectiveness and efficiency of learning, while accommodating different learning styles. Potential advantages such as providing consistency in educational approach, increasing access to information for those unable to attend formal courses and presenting an educational approach philosophically consistent with liberal curricula were identified.

Aims

- To develop and formatively evaluate the multimedia program, 'A Mentor's Guide to Assessing Student Nurses', that facilitates the acquisition of skills in the observational assessment of student nurses' performance.
- To evaluate the educational effectiveness of this program through comparing the multimedia method with the traditional teaching method in a randomised controlled trial

Development methods

The program was constructed using behaviourist, cognitivist and constructivist methods in three main conceptual areas, 'The Assessment Process', 'Demonstration of Assessment Concepts', and 'Assessment Simulation'. A strategy of formative evaluation guided the instructional design. A questionnaire was developed to indicate 'assessment potential' in mentors through measuring their anxiety, attitude, and self-efficacy in relation to the assessment of student nurses.

Evaluation methods

The questionnaire was found to have acceptable validity, internal consistency and test re-test reliability and was used for pre and post measurement in a Randomised Controlled Trial. Both experimental and control interventions were carried out for groups of 6-10 subjects. Mentors (n=130) who agreed to take part were randomised individually to experimental (n=62) or control interventions (n=68) which were carried out as 7 control groups and 9 experimental groups over 6 months. Mentors were recruited on a voluntary basis and randomly allocated to study groups. Baseline data was collected immediately before the educational intervention, then immediately after, with follow-up telephone interviews of a purposive sample (10%, n=12) within 8 weeks of the trial ending. This comprised one general question followed by three questions on anxiety, attitude and confidence in relation to the subject's assessment of students' clinical performance. The constitution of study groups was compared for systematic differences using the Chi-

square test; baseline scores for anxiety, attitude and self-efficacy were compared using Mann-Whitneys' U Test; improvements in anxiety, attitude and confidence between baseline and immediately after the intervention were compared within groups using the Sign Test, and between groups using Kruskal-Wallis' one-way ANOVA.

Evaluation results

Mentors recruited to the study were representative of Scotland in terms of their place of employment, gender, and clinical grade. Years since qualification, experience of assessment, educational background, professional qualifications, and frequency of exposure to students in the workplace, scores for anxiety, attitude and self-efficacy, were all similar between computer and control intervention groups at baseline.

The multimedia program was no more effective than traditional teaching in reducing anxiety or improving self-efficacy but attitudes in the experimental group improved more than controls. The telephone interviews indicated that cognitive conflict may have had a confounding effect: the disconcerting effect of being challenged on long-held beliefs increased anxiety in some subjects, negating the reduction in anxiety that was anticipated. These findings in the experimental group differed from the heightened anxiety that was experienced in some subjects in the control group, who expressed frustration and disappointment at not having had their learning needs met.

Discussion

The 'novelty value' of the computer program may have positively influenced the effectiveness of the educational intervention for the experimental group. This was argued to be a motivational factor integral to multimedia, distinguishing it from traditional teaching methods. The failure of self–efficacy improvement to reach statistical significance between groups was considered a function of instrument design and small sample size.

It was concluded on the basis of the questionnaire results and telephone interview findings that the program improved mentors' assessment potential, most evident in the statistically significant improvement in attitudes. The research design was satisfactory in providing a representative sample to study and the 'assessment potential' questionnaire was user-friendly, producing data of value. In developing the computer program, the inexact science of performance assessment and the difficult nature of mentoring were affirmed. Problems with clinical assessment were difficult to completely resolve, but it was possible to achieve a balanced approach based on the available literature.

The computer package developed and tested in this study shows that assessment can be effectively taught to mentors using a 'stand-alone' package and that this approach should be included in the range of methods used for teaching mentors how to assess. This is different to computer-mediated communication that typifies networked applications, where learning is a function of the vagaries of the student - teacher relationship. In stand-alone programs, an instructional design can be created with proven capability in achieving high levels of human-computer interaction that can deliver repeated quality assured learning in a way that computer-mediated communication approaches couldn't. There is potential for convergence, achieving the best of both worlds: computer-mediated communications courses have been introducing more 'e-tivities' in a departure from sole reliance on dialogue and stand-alone applications can link with readily updated web-pages and web-tools such as discussion groups. The second approach was taken with the latest version of 'A Mentor's Guide to Assessing Student Nurses'.

Recommendations

Nurse Education

- Implementation of 'A Mentor's Guide to Assessing Student Nurses' now by marketing it to mentors and lecturers, packaging as part of a level three Scotcats module, and using it as the basis of a UKCC course for mentors on assessing students.

- Encourage nurse educators to consider CAL in their portfolio of approaches to learning, perhaps as an optional, student centred, fixed resource in PBL curricula.
- Partnerships between media publishing companies and educational institutions ought to be developed to increase the availability of high quality learning resources.

Research

- Conversion of this program to web pages, with streaming video ought to be explored.
- Methods of integrating nursing thesaurus and language classification systems into CAL ought to be explored. This would allow the interpretation of free-text entry in order that higher levels of human-computer interaction can occur.
- The relationship between state anxiety and cognitive conflict needs to be explored.

Nurse Management

- A Liberal approach to in-service education ought to be adopted, and the use of student centred methods such as CAL encouraged.
- Practising nurses ought to be encouraged to capture authentic clinical media to create a database for use within in-service and university education
- Knowledge of programming, multimedia, and instructional design should be developed in key clinical staff to enable nurses to lead the development of nursing software.

Acknowledgements

For
Jacqueline, Joanna, Claire and Liam

I would like to take the opportunity of thanking those who enabled this work to develop and progress beyond the planning stage. The most influential have been

Margaret Smith, then Director of Nursing and Quality, Victoria Infirmary NHS Trust.
Miss C.A Asher, OBE, then Principal of Glasgow College of Nursing and Midwifery.
Professor Barbara Parfitt, Head of Department, Nursing and Community Health, Glasgow Caledonian University.

Media production would not have been possible without the contribution of

The patients and staff at Victoria Infirmary NHS Trust.

Friends at the former Glasgow College of Nursing and Midwifery, especially Jacqui McKechnie, Jan Dickie and Mary Stott.

Students of the class of 1992 and 1994,

Bill, Catriona, Les and Drew

Trevor Grant (voice over).

For the guidance and supervision received, my thanks to

Dr Ray Jones, Senior Lecturer, Department of Public Health
Professor Edith Hillan, Nursing and Midwifery School

For financial support, my thanks to

The National Board for Nursing, Midwifery, and Health Visiting for Scotland

For evaluation, my thanks to all who contributed their particular expertise, especially

Kathy Duffy, Ellen Robertson, Margaret White, Yvonne Robb, Jeanette Brown, Eugene Quirk, Erica McAteer, and Robin Shaw. In addition, the prospective mentors who were involved in the Randomised Controlled Trial are thanked, all 130 of them.

Finally for the task of digitising and integrating media into Authorware, a special acknowledgement and thank you to Kathy Trinder, multimedia designer and author, for her invaluable expertise.

Declaration

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Glasgow, Faculty of Medicine, Nursing and Midwifery School, and Department of Public Health. Unless stated otherwise, the work is that of the author.

Table of Contents

List of Appendices

List of Tables

List of Figures

List of Abbreviations

ALHR	Four digit number representing accuracy, leniency, harshness and range, calculated from the user's input to the simulation, which forms the basis for feedback.
ANOVA	Analysis of Variance
CAI	Computer Assisted Instruction
CAL	Computer Assisted Learning
CBI	Computer Based Instruction
CD-I	Compact Disc Interactive
CD-ROM	Compact Disc, Read Only Memory
DOF	Degrees of Freedom
DVD	Digitally Versatile Disc
ENB	English National Board
HCI	Human Computer Interaction
ICT	Information and Communications Technology
IVD	Interactive Video Disc
NBS	National Board for Nursing, Midwifery and Health Visiting for Scotland
NHS	National Health Service
NVQ	National Vocational Qualification
PBL	Problem Based Learning
RCT	Randomised Controlled Trial
SAI	State Anxiety Inventory
SD	Standard Deviation
SPSS™	Statistical Package for the Social Sciences
STAI	State-Trait Anxiety Inventory
SVQ	Scottish Vocational Qualification
TLTP	Teaching and Learning Technology Project
UKCC	United Kingdom Central Council for Nursing, Midwifery, and Health Visiting

Chapter 1: Introduction and Study Background

1.1 Summary

This chapter introduces and provides a concise rationale for the development and evaluation of a multimedia program to help mentors assess the clinical performance of student nurses. The effects of major organisational change were a consideration throughout the study, exacerbating perennial concerns over the validity and reliability of student nurses' clinical assessments, while obscuring medium and long-term views of the future of nurse education. Meanwhile, a revolution in learning technology was underway, offering multimedia as a potential solution to the logistical difficulties of providing a quality preparation for those responsible for assessing students: their mentors.

1.2 Study background

Evaluative research, as a logical and systematic investigation to assess the effects of a specific programme, practice or method [1], was chosen to evaluate the enclosed program, 'A Mentor's Guide to Assessing Student Nurses' (appendix 1). This type of research can be impeded by problems of an interpersonal, organisational or political nature, implying that more than research skills are required [2]. Such problems occurred during the present study, as designing, constructing and evaluating a multimedia program took place during a period of uncertainty caused by unprecedented organisational and educational change.

A period of nurse education reform, followed several policy directives from the United Kingdom Central Council for Nursing Midwifery, and Health visiting [3-5]. The outcome of this became known as 'Project 2000'. Nursing courses were re-designed to provide both a professional qualification and a higher academic award, beginning at diploma level. In addition, colleges of nursing decreased in number and integrated with institutions of higher education, resulting in a much higher rate of organisational change than had previously been witnessed. During this period, many things changed, but the fundamentals of preparing nurses remained: courses were designed to develop students' practice skills in addition to theoretical knowledge. As such, fifty percent of time in pre-registration study was spent on placement. This was in the tradition of many professions

already in higher education, such as teachers, doctors and social workers, but what became contentious was the validity of this work-based learning and how to accredit the 'swampy lowlands' of practice [6] (p43) within academic programmes [7]. As a distinction developed between being 'academically able' (at diploma, or degree level), and being 'clinically competent' for registration, some institutions began to find ways of crediting nursing practice with academic merit [8, 9]. However, Erault and Cole [10], in a large study of the assessment of competence in the professions found that assessment and verification procedures were poorly developed for workplace based assessment. For such reasons, the National Board for Nursing, Midwifery, and Health visiting for Scotland (NBS) insisted that the assessment of practice be as reliable and valid as the assessment of theory [11]. One major difficulty in achieving this has been the inadequate preparation of registered nurse 'mentors' to support and supervise students on clinical placement [12-14]. In some areas of the United Kingdom, mentors have the additional responsibility of assessing students' practice. Quality control mechanisms in the selection, preparation and support for nurses given this responsibility would seem to be essential, but the literature has shown that this has not always been the case [14-18]. This situation provided an imperative to find methods and means of developing mentors' skills and knowledge to enable them to support students. This provided the initial motivation for undertaking this study.

In addition to changes in education, rapid growth in communications technology provided the backdrop to this study, from 1995 to 2001. During this time, interactive multimedia was hailed as the answer to training and education needs [19], with consequences for both providers and recipients of higher education [20, 21]. A computer-based interactive multimedia program seemed an attractive option for the preparation of mentors, for several reasons: it could be distributed easily, and, if constructed according to educational principles, could provide a quality educational experience for users. This research project has involved the instructional design, development, and evaluation of such a program. Through quality mechanisms during production, a program evolved that could be used without ethical constraint in a randomised-controlled trial, in order that it could be compared with traditional teaching.

Coping with enormous change has been a challenge to the management of this project. However, the experience has reinforced the researcher's belief that learning technology, such as at the core of this study, is crucial to evaluate to determine if, and to what extent, educational objectives have been achieved. The growing number of practising nurses and other professionals investing in programmes of learning, both at a distance and in the workplace require the re-assurance that learning technologies are appropriate and effective [22-24].

1.3 Conclusion

A custom-built program 'A Mentor's Guide to Assessing Student Nurses' was developed to prepare new mentors to assess students. To achieve this it was necessary to demonstrate that the program was capable of delivering learning outcomes at least as effectively as traditional means. Through an extensive literature review, program development and evaluation, this study provides an insight into the process of learning using computers, and adds to the body of knowledge on learning technology and the assessment of clinical practice.

Chapter 2: Organisation of the Literature Review

2.1 Summary

This chapter outlines the approach taken to reviewing the literature, describing the strategy for coping with a continually growing, potentially overwhelming volume of literature ranging from the anecdotal to major research projects, in the key areas of

- Assessing Student Nurses' Clinical Performance: Context, Theory, and Practice
- The Observational Assessment of Clinical Performance
- Mentorship
- Learning Technology and its Evaluation

2.2 Scope of the literature review

The literature review served as an orientation to what was known of the subject, and created a conceptual context for this research. Four activities involved in reviewing the literature, identified by Polit and Hungler[2] were used to provide structure. These were

- defining the scope of the literature review,
- identifying sources of information,
- reading and reviewing the literature, and
- constructing the written review.

This review firstly analysed the theory and practice of student nurses' performance assessment, particularly within the post-Project 2000 educational context. Secondly, the relationships between the assessor role and other aspects of mentorship were explored. Finally, aspects of learning technology as they related to the instructional design and evaluation of the program, 'A Mentor's Guide to Assessing Student Nurses' were analysed. This wide-ranging literature review provided direction and a theoretical basis for the construction of the program and its evaluation. As research in these fields can be found in the United Kingdom, Europe, Australia, as well as in North America, inclusion has not been geographically or culturally defined. However, only literature written or translated into English was sought.

A variety of sources were accessed, such as traditional and electronic journals, books and reports. Sources were primarily from nursing research, nurse education, and general education, although other subject areas such as psychology, management and computing were searched for literature of relevance.

2.3 The organisation and process of conducting the literature review

The volume of literature on assessment, learning technology and mentorship has grown to overwhelming proportions at an exponential rate. However, limiting searches to the most recent and relevant publications helped narrow the field considerably. The purpose of this review was to provide a succinct account of essential concepts and their origins. Primary sources were sought from 1990 onwards, although earlier publications have been utilised, especially frequently cited, or seminal work.

2.4 Reviewing the literature

Only the first three of five categories of literature identified by Polit and Hungler [2] were used in this literature review, because of the highly subjective nature of the others. These were facts, statistics and findings; theory or interpretation; methods and procedures. Initial scanning of publications allowed poorer quality literature to be discarded. This critical review process was designed to encompass a wide range of sources to ensure that potentially useful information was not omitted. The quality of studies and potential generalisability of findings therefore varied.

2.5 Organisation of the written review

The main topic areas provided a logical means of dividing and structuring this review. These were:

Chapter3 Assessing Student Nurses' Clinical Performance: Context, Theory, and Practice.

Chapter 4 The Observational Assessment of Clinical Performance.

Chapter 5 Mentorship.

Chapter 6 Learning Technology and its Evaluation.

Chapter 3: Assessing Student Nurses' Clinical Performance: Context, Theory, and Practice

3.1 Summary

This chapter, in exploring the assessment of student nurses' clinical performance, explores the links between teaching, learning and assessment, and the importance of philosophical consistency within the curriculum as a whole. The influence of the theories of Benner in relation to domains of nursing practice and skills acquisition; and Kolb, Steinaker and Bell in relation to experiential learning has been examined. The educational context post Project 2000 has been described, and its impact evaluated. On balance, the evidence indicates that this liberal approach to the education of nurses has been mainly beneficial. However, prevailing traditional philosophies have caused conflict, especially between the teaching and assessment of theory and the teaching and assessment of practice. In addition, clinical assessment caused difficulties through lack of consistency and poor reliability of rating, instructor subjectivity, questionable validity of assessment tools, poorly utilised assessment criteria and reluctance of assessors to negatively rate or fail students. This can be attributed partly to the difficulties inherent in clinical assessment, and partly to the poor preparation of assessors. Further analysis of the assessment of clinical practice identified contradictory and confusing terminology, such as competence and performance. These concepts are explored, identifying their differences, and their respective strengths and weaknesses. The chapter concludes that clinical 'performance' assessed in the practice context, is more appropriate than the assessment of competence, which can be undertaken in a number of ways. Benner's Domains of Nursing Practice were identified as key 'aspects of performance' that, being capable of independent assessment, represented a compromise between atomistic and holistic approaches to assessment.

3.2 The educational context of this study

3.2.1 Project 2000

The purpose of Project 2000 was to prepare nurses to meet health care needs in the 1990s and beyond [4]. Its' philosophy originated much earlier, in reports such as those by

Horder and Briggs [25, 26]. The key element was that nursing students should be treated as supernumerary whilst learning in practice [4] because until then, students were employees in vocational training schemes, paid to care for patients. In 1989, permission for thirteen demonstration sites was given and in 1992, new courses commenced throughout the UK [27]. These courses embodied a number of important principles, identified by May et al [28]. Students were learners rather than employees and teaching was through student centred methods. Nursing theory and practice were integrated and based on research; clinical experience was education-led, and programmes had a 'health' as opposed to an 'illness' orientation, in a departure from the 'medical model'. Finally, students were being prepared as lifelong learners and the assessment of practice was competency-based [28]. Such major changes precipitated a steady flow of publications reporting on its progress [18, 28-31].

3.2.2 The success of Project 2000

Many of the initial studies of Project 2000 focused on its introduction [31]. Common themes were the inadequate preparation of clinical staff, heavy workloads, and low staffing levels having an adverse effect [28, 30]. Despite this, there were many accounts of its successful introduction [18, 32, 33]. Later studies reported on established Project 2000 courses rather than initial, often transient experiences. One example from a number of similar studies surveyed students in their final year [31]. These students (n=131) expressed a high level of satisfaction with Project 2000 in terms of value, breadth and relevance. Some aspects of the course were assessed as poor, for example, organisation and liveliness. Practice was perceived as having higher value than theory, however, on balance, there was general satisfaction with the intellectual demands of the course. This contrasted with the findings of a study by French [34] cited by Parker and Carlisle [31]. In this study, students considered their course to be intellectually challenging, while lacking an emphasis on practice. This became a recurring theme in subsequent literature, suggesting that Project 2000 courses produced nurses that perhaps understood the context within which they practised, but were not necessarily competent to practice [35].

3.2.3 Satisfaction with the product of Project 2000 courses

Several studies reported deficiencies in the practical and organisational skills of students at the point of qualification [36, 37]. Such findings have been reinforced by the perceptions of nursing management [38]. However, the validity of the views of managers can be questioned, given that their expectations of newly qualified nurses in the UK have been found to be unrealistically high [27, 39]. Studies in the United States have had similar findings. For example, one research study reported significant differences between the Baccalaureate degree competencies that directors of nursing (n=142) expected, compared to those that were witnessed [40]. In addition to unrealistic expectations putting pressure on students, there have been extensive changes to the organisation and delivery of health care affecting the learning environment of many placements, which were not as secure and supportive of students as they once were. There have been no research studies to identify clear cause and effect relationships in this situation. However, one study indicated that any deficiencies were short lived. Nurses have been reported to be more comfortable and effective in their professional roles within six to nine months of qualifying [36]. Another study by Gerrish [41]. compared separate groups of qualified nurses, the first group seven years before Project 2000 (n=10), the second group six years following its introduction (n=25). Findings reveal that although newly qualified nurses in 1998 felt just as inadequate as those in 1985, they developed a more active style of learning and found the transition from student to professional nurse less stressful. This demonstrates that, at least for some, Project 2000 courses may have been advantageous.

The Project 2000 theory-practice gap

One explanation proposed for newly qualified nurses not meeting the expectations of service staff has been that Project 2000 has widened the 'gap' between practice and theory [42]. This gap, minimal in vocational training [43, 44], has reportedly been increased by Project 2000 [27, 29, 45]. This has been attributed to the programme being educationally led and underpinned with principles of adult learning; nursing students being taught without having relevance to practice explained [8], and to an increase in classroom based learning [27]. More pointedly, it has been speculated that the 'gap' is the consequence of

poorly prepared mentors [18] who failed to link nursing practice to theoretical concepts. It may well be a combination of these factors.

Wastage rates in Project 2000

An easily measured performance indicator of the success of nursing courses has been completion and wastage rates. Projected improvements as a consequence of introducing Project 2000 [4] have been confirmed, as average attrition rates improved from 35% in 1980 to around 15% nationally in England and Wales in 1996 [46, 47].

Despite its many critics, studies have generally concluded that Project 2000 programmes have been successfully implemented and have continued to develop [12, 31, 36, 37]. Overall, Project 2000 courses seem capable of preparing students to become qualified nurses, as summed up by Elkan, who, following an in-depth review of literature, concluded that the main value of Project 2000 has been to provide a framework within which high-quality nurse education could occur [30].

3.2.6 Intractable difficulties with Project 2000 courses

The generally positive findings in relation to Project 2000 did not give rise to complacency. One recurring criticism focused on clinical assessment, a long-standing problem in nurse education [13, 48]. Problems have included lack of consistency and poor reliability of ratings; instructor subjectivity and bias; questionable validity of assessment tools; poorly utilised assessment criteria; reluctance of assessors to record negative ratings, and an unwillingness to fail students who should not pass [13, 49-53]. Factors thought to contribute have been the complexity of the clinical environment, the conflict arising through students providing care for patients while learning, the difficulties inherent in the process of evaluation, and the lack of uniform, agreed standards. In addition, there have been few well-controlled studies and a dearth of evidence about the nature of clinical learning[54]. The next section will explore why these difficulties in assessing students should be of continuing concern to nursing, and why the search for a valid and reliable method has, for some, become a search for the 'Holy Grail'[50].

3.3 The purpose and importance of the assessment of student nurses' practice

3.3.1 Introduction.

Despite the strategic position that assessment holds within all education systems [55], Peach et al [39] found that in nursing practice, assessment strategies have not been effective in identifying poor performers. Learning outcomes have been vaguely stated, assessment documents unclear, and assessors ill prepared and lacking in feedback with regard to their own performance. This chapter provides an indication as to why the assessment of nursing practice has been the source of so much angst.

3.3.2 Definitions and terminology

The terms, 'assessment' and 'evaluation', have been used interchangeably. However Burnard made a distinction, in which to assess was to identify a particular state at a particular time, and to evaluate was to identify the success or otherwise of a course of action [56]. Assessment thus occurs at the outset of an educational experience and evaluation at the end. As evaluation leads to re-assessment, a cycle of assessment, evaluation and re-assessment develops. However, North American publications have used the term 'clinical evaluation', while the United Kingdom have used 'clinical assessment' to refer to the same process. This has been further complicated by the use of similar methods and tools for both assessment and evaluation [56]. In keeping with current practise within the UK, the term 'assessment' has been used in this study.

Assessing a student has often been considered a process of obtaining some measure of their ability to integrate cognitive, affective, and psychomotor skills in the delivery of nursing care [57, 58]. Some have perceived assessment to be a diagnostic or feedback mechanism essential to planning further learning activities with the individual, in a constructivist tradition [59-62]. However, a range of subtle, conflicting functions have also been shown to exist [63, 64]. For example, it has been a means of maintaining the status quo in nursing, using the power of authority and the threat of imposing a poor assessment on those who dissent. Ironically, this has often been while socialising students to the norms and values of a caring profession. Conversely, for some students, assessment has hedonic value in the acclaim received for relieving the professional nurses of part of their

workload. Assessment can thus be a powerful motivating factor for students, where overt and covert reinforcement has an influence, in the behaviourist tradition [65]. The assessor may also serve as professional 'gate-keeper' by determining who progresses to become registered practitioners. Finally, there has been a bureaucratic function of clinical assessment, which can theoretically provide a differentiation between Diploma, Degree and Masters students [9, 38, 66, 67]. However, clinical assessments have failed to make this distinction, much to the concern and frustration of those within the profession [48].

This discussion of the different functions of assessment has revealed a source of potential conflict. Conflict in the minds of the assessor and the student, conflict between assessors and conflict between lecturers, students and assessors. An additional source of conflict has emerged in Project 2000. The philosophy underpinning a traditional 'competence' approach to clinical assessment is more in keeping with NVQ training programmes [68], than the liberal, holistic approach espoused in Project 2000 curricula. An objective of curriculum development and course planning is to create a harmonious relationship between teaching, learning and assessment. If the educational principles underpinning the curriculum imply active, holistic and contextual learning [69], then so ought the process of assessment.

Morle [70] argued that how assessment was approached depended upon how the individual viewed its purpose: whether it was primarily a process for growth or a means of control. Those preferring the control aspect were inclined towards basing their assessment on rules, regulations and conformity, forcing students to focus on surviving the placement and achieving a pass rather than developing skills and knowledge that would contribute to their professional development. One of the objectives of 'A Mentor's Guide to Assessing Students', has therefore been to focus clearly on an approach to assessment consistent with a liberal educational philosophy. Another has been to help prospective mentors achieve a shared understanding of the purpose of assessment in the practice setting.

3.3.3 Professional accountability

Registered nurses must be satisfied, contractually and professionally, that students in their care can safely undertake delegated nursing activities, for which they can be held accountable [71, 72]. This extends to ensuring that students continue their professional development, and have this confirmed through assessment. There have been indications that this has been neglected by some mentors. For example, avoiding the unpleasant task of failing students has been a frequent finding in the literature [13, 48, 52, 53, 73].

3.3.4 Conclusion

Sound reasons have existed for assessing student nurses accurately and appropriately, such as to determine whether student nurses' educational needs have been met and to maintain standards. To be valid, an assessment should measure what has been learned, but what has been learned by students can be determined by what will be assessed[55, 74]. In this way, assessment can drive learning[75-77]. Rowntree [55] therefore advised that learning and assessment should be considered together. How students learn on placement is therefore considered next before analysing the assessment of practice in more detail.

3.4 Learning in and from practice

3.4.1 Introduction

Liberal and traditional educational philosophies have both had their place in nursing curricula. Whereas a liberal, student centred approach to education has been used within many Project 2000 curricula, the UKCC has indicated that the assessment of theory ought to be rigorous, and competency based, hinting at a somewhat more traditional approach.

3.4.2 Conflicting traditional and liberal educational philosophies

The history of Western education has demonstrated that the roles of teacher and student have varied depending on whether the prevailing view had been 'traditional' or 'liberal' [78]. Education in the 'traditional' sense has often been a didactic, one-way process, undertaken in a formal educational or training institution, the teacher being the knowledgeable expert imparting knowledge and skills to students [79]. Such an approach

to education has been acceptable at certain times, in certain cultures and occupations. Eraut [80] argued that professional expertise in such traditional curricula is identified with the possession of highly theoretical propositional knowledge, which contrasts with liberal curricula, where knowledge and expertise have been identified with learning that stems from experience. Liberal, student centred methods in nursing, such as problem based learning, reflection, tutorials, student-led seminars, simulations, small group work and educationally-led practice placements, have thus allowed students to take the first steps towards acquiring this professional knowledge and expertise [77, 81-83].

Maynard [84] in a longitudinal study of nursing graduates identified the experiential component of the course as the essential factor influencing the development of competence and critical thinking, often considered the essential components of professional education[85-87], whereas individuals educated by traditional didactic methods have found difficulty developing thse skills. Liberal, student centred principles of adult learning have therefore been considered the most appropriate approach to the education of professional nurses [45].

3.5 Theoretical perspectives on learning from practice

This section examines the main theoretical influences on nurse education, with particular reference to learning in practice settings, comparing the work of Benner, Kolb, and Steinaker and Bell. Consequences for the assessment of practice have been identified.

3.5.1 Patricia Benner[88]

Benner described a taxonomy of levels of practice in nursing based on a model of skills acquisition detailed in research by Dreyfus and Dreyfus [88, 89]. One hundred and thirty adult and paediatric intensive care nurses were interviewed in groups to determine whether this model could be applied to nursing. Narratives were used as a method of analysing nursing practice, and of demonstrating how novice nurses became advanced beginners, developed competence, became proficient and then developed expertise (table 1). It was found that clinical knowledge was gained over time and was frequently an unconscious process that involved passing through identifiable stages [88]. This confirmed

that the Dreyfus and Dreyfus model was applicable to nursing [90]. Benner's work has been criticised from a positivist perspective [91, 92]. Concern centred on the prominence given to clinical judgement and intuition in the process of developing expertise, said to undermine the value of propositional knowledge and formal theory. Her work has also been criticised for its focus on intensive care and no research had determined whether it had validity in other areas of nursing, such as pre-registration education. Benner has been described as a promoter of oppressive traditionalism [93] while others have described her work as empowering, enthusing, and challenging [94].

Table 1: Benner's 'Novice to Expert' model [88]

NOVICE
ADVANCED BEGINNER
COMPETENT PRACTITIONER
ADVANCED PRACTITIONER
EXPERT

It has been argued that specific skills characterise Benner's levels, and that nurses may be placed within this hierarchy based on their clinical performance. In order to undertake her research, Benner identified seven 'Domains of Nursing Practice' as shown in table 2. These were used as a means of classifying skills and competencies identified through work-role analysis of one thousand, two hundred practising nurses, using self-reports and selective interviews [88].

Table 2: Benner's Domains of Nursing Practice [88]

THE HELPING ROLE
TEACHING-COACHING FUNCTION
DIAGNOSTIC AND PATIENT MONITORING FUNCTION
ADMINISTERING AND MONITORING THERAPEUTIC REGIMES
MONITORING AND ENSURING THE QUALITY OF HEALTH CARE PRACTICES
EFFECTIVE MANAGEMENT OF RAPIDLY CHANGING SITUATIONS
WORK-ROLE COMPETENCIES

Despite criticisms, this framework, together with her model of skills acquisition have been used widely in nurse education [48, 66, 91, 95-100]. Strengths of Benner's work have been its intuitive appeal to nurses, its flexibility and its applicability across a wide range of settings, and its usefulness in categorising and assessing competence [48, 92]. The 'domains of nursing practice' as a framework for categorising different aspects of a student's performance thus has validity. This has therefore been adapted for use within 'A Mentor's Guide to Assessing Student Nurses'. Unfortunately, Benner's taxonomy of Novice-Expert (table1) would be inappropriate to use to describe students' development in undergraduate nurse education, as the goal has been to develop competence, rather than expertise [39]. A more appropriate 'model' to describe the transition that occurs in nursing undergraduate students was therefore required, that was focused on practice. Students have many experiences while in practice, and are expected to learn from these. Such experiential learning has been described by Burnard [101] as any learning that increases the facilitation of experiential knowledge, that is, knowledge revealed through practice. An analysis of two theories of experiential learning that have been influential in undergraduate nurse education now follows.

3.5.2 Kolb's theory of experiential learning [102]

Of the experiential learning models, Kolb's is best known. He argued that learning extended beyond the classroom to become a life-long, holistic, active process linking education, work and personal development. Kolb's educational perspective parallels the developmental theories of learning such as Piaget [103] on child development, and Erikson [104] on social development. Kolb's Learning Cycle model places 'experience' central to the learning process as the basis for observation and reflection (Appendix 2). This model has provided a useful, comprehensive and systematic statement of the theory of experiential learning. The impact of experiential learning theory on nursing curricula has been pervasive. In a survey by Burnard [81], it was found that most student nurses had engaged in college based experiential learning activities such as small group work (98%), role play (96%) and activities that encouraged the development of clinical skills (94%). Kolb's theory provided the justification for much of the experiential learning undertaken by student nurses, but the appropriateness and theoretical basis of many methods of

experiential learning have not been universally accepted [45, 105]. In addition, Kolb's approach appears to have been overshadowed by Steinaker and Bell's [106] experiential taxonomy as a pragmatic framework for assessing students' clinical performance [66].

3.5.3 Steinaker and Bell's experiential taxonomy [106]

Steinaker and Bell argued strongly that human experience could not be classified into cognitive, affective and psychomotor 'learning domains' as described in Bloom's taxonomy [107]. Neither should 'experience' be an entirely separate domain as proposed by Sheahan [108]. Instead, 'experience' was argued to be integrative and capable of simultaneously addressing two or more learning domains. Steinaker and Bell thus produced a taxonomy of experiential categories, through which learning develops.

Table 3: Steinaker and Bell's experiential taxonomy [106]

EXPOSURE
PARTICIPATION
IDENTIFICATION
INTERNALISATION
DISSEMINATION

Each of these stages has been described according to student behaviours, thoughts and feelings. For example, on first encounter or "exposure" to an experience, the student may not physically contribute to the activity but uses senses of perception and mental processes in attempting to understand and theorise about ways of contributing. The teacher's role is to explain, encourage and motivate [106].

Brooke et al [109] found this experiential taxonomy useful in demonstrating student progression, from being a passive attender at 'exposure', to becoming an influencer at 'dissemination'. It has been used widely in clinical assessment tools [48, 66, 110]. Evidence of construct validity was provided in one study where a close relationship between assessments made with the experiential taxonomy was found with a tool used by students

to record their personal learning [111]. In addition to being a useful taxonomy for assessing practice, it has also been used as a conceptual framework for curricula [106, 109]. This taxonomy of levels would appear to be much more appropriate to undergraduate education than Benner's Novice to Expert taxonomy (table 1). Because of its popular appeal [48] Steinaker and Bell's taxonomy has been adapted for inclusion within the program 'A Mentor's Guide to Assessing Student Nurses' (appendix 1), to provide a useful method of distinguishing between the different levels of performance that students exhibit.

3.6 The relationship between context, learning and assessment

Within placements, learning opportunities differ because of the large number of variables present, such as patient dependencies, patient illnesses, nursing needs, staffing levels, skill mix, and personalities [49, 51]. Hepworth [63] asserted that as nursing is grounded in the uniqueness of every new nurse/patient relationship, then the assessor requires to judge students' performances in the light of their own perception of each situation. This has been complicated by the suggestion that information available to students, upon which actions are based, is different to the information available to assessors whose perspective is different. This could have a significant effect on the assessment of students' performance [112].

To accommodate such diversity in learning environments, Orchard [113] argued that appropriate student performance expectations should be developed to meet specific clinical settings. These expectations ought to be reasonable, be applied consistently and be communicated to students before their implementation. One clear advantage of this would be that students' assessments would be specific to the context of learning.

An issue raised in relation to quality assurance of work-based learning in higher education has been the vagueness of learning outcomes[114]. However, this can be a distinct advantage if it allows interpretation in different contexts and is consistent with adult learning and student centred approaches, such as problem-based learning. A more flexible, situated approach to assessment of practice within higher education has been

argued by Winter [114] to be the most logical result. The UKCC, by contrast, has become more concerned than ever with the assessment of student nurses' practice,[8, 9, 11, 115] with some calling for centralised control [48].

3.7 Conclusion

This section has focussed on the theories of learning as applied within practice situations, and some of the influences upon this. The works of Benner and Steinaker and Bell were found to be of particular relevance, and the multimedia program has been designed to include these: the Benner framework as a means of grouping competencies, and Steinaker and Bell as a framework for judging their achievement. However, as the program ought to encompass the process of clinical assessment, from beginning to end, contentious issues in relation to this now follow.

3.8 Clinical assessments: what are they a measure of?

3.8.1 Introduction

This section develops a discussion on different perspectives of clinical assessment and clarifies contradictory and confusing terminology, such as competence, competency, capability and performance. These were important to consider before choosing an appropriate instructional design for the multimedia program.

3.8.2 The difficulties associated with clinical assessments

Studying the history of professional nursing reveals many attempts to analyse and rationalise the methods and processes of student nurses' clinical assessments. Frustration has been expressed at the inability to resolve the inherent difficulties that have been encountered [50, 51, 63, 72, 113, 116, 117]. Experimentation with assessment has occurred over the decades, together with a tendency to change from one approach to another without systematic evaluation [48]. These difficulties have not been unique to nursing; professions such as medicine and occupational therapy have had similar concerns [59, 118]. It has been argued that the fundamental problem has been trying to treat objectively, a phenomenon that is inherently subjective [119, 120], and that intuitive thinking is a necessary part of the assessment process and ought to be accommodated [63, 121]. In light of this, and as the

UKCC has adopted a competence approach to curricula, including the assessment of practice, the consequences are discussed next.

3.8.3 Exploring the concept 'competence'

Crucial to carrying out any assessment process is deciding what should be assessed [122]. Competencies have been the focus of clinical assessments in post- Project 2000 curricula [11, 12]. Therefore, the concept 'competence' and its different interpretations have been explored.

3.8.4 Definitions of competence

The term 'competence' is a semantic label assumed to be a general attribute of professionals [123]. Various definitions exist, most being concerned with the ability to perform skills or activities within an occupation[124, 125], however, it is a multidimensional construct open to different interpretations [9, 48, 72, 126-128]. The Commission for Nursing and Midwifery Education qualified this simple definition, by emphasising safety, and the absence of direct supervision in the performance of activities [39].

The capacity to conceptualise competence in different ways has caused much confusion. It has been claimed to represent both a basic level of functioning [129, 130] and a higher level of functioning[131]. Clearly a gulf exists between these perspectives. No single definition has been found to satisfy all requirements [9, 132]. Norman et al [48] supported a definition indicating that having the potential to carry out an activity correctly ought to result in an individual being judged competent in that regard. Competence, in this sense, is based on how people approach their work rather than whether they have been judged in the performance of it. This has been argued to accommodate intuition and tacit knowledge [80] and can be assessed in a variety of indirect ways, but its weakness is that it does not necessarily involve the observational assessment of performance. It has been argued that having the capacity to do a task effectively and knowing how to do it does not necessarily mean that it will be effectively accomplished [58]. The proposition that any

measure of competence ought to focus on actual performance, rather than the individual's capacity or capability to be realised at some future date, has found support[58, 133].

3.8.5 Exploring the concept 'performance'

Performance, like competence, has problems of definition [134-137]. It can be distinguished from competence and capability by focusing on what is actually done in day-to-day practice [118] and is in keeping with the broader term 'competency'[58]. However, a further distinction has been made. Competence may be determined through assessment methods such as reflection and journals, whereas performance may only be determined by an assessor observing the student in the actual situation [133]. Performance based evaluation of a student nurse in the clinical area, however, has created difficulties because

- as it is based on observation it is open to bias and subjectivity [48, 49, 138]
- the nature of the clinical setting is critical to the student's performance, where different personalities and patient conditions must be taken into consideration [49, 51]
- the assessment is inevitably based on a sample of the student's total experience [49, 138]. Unfairness and inaccuracy can result, as only a small part of total performance is ever available for consideration [139].
- the student requires to adjust to the real life situation and apply what has been learned in class and lab situations, while being assessed. There has been a consensus that this is undesirable, because the distinction between teaching and learning activities, and formal assessment is unclear [49, 51, 117, 140-142].
- Learning experiences for students in different placements or at different times have been difficult to equate [49]

For these reasons, the assessment of nursing practice has often involved a mixed approach, perhaps combining observational assessment of performance with an assessment of capability, for example, through keeping a reflective diary.

3.8.6 Competence and the UKCC

McGaghie [123] argued that the meaning of competence for any profession should begin by a description of its domains, boundaries, and relations. This has arguably been undertaken for nursing by Benner [88] in identifying the 'domains of nursing practice' through her research (table 2). This perhaps contrasts with the statutory competencies for Nurses, Midwives and Health Visitors that were introduced in rules 18 and 33 in 1983 [143], which were updated as Rule 18a immediately prior to the implementation of Project 2000 [27]. The identification of these competencies, although claimed to have produced a more coherent approach to assessment [127], have been criticised for being generalised, duplicated, unclear and lacking in any order [144]. This does not comply with criteria of clearly identifying the boundaries of competence, nor does it allow the desirable assessment of different levels of competence [48, 123].

Many institutions used these arguably flawed UKCC competencies as the basis of their nursing and midwifery curricula, causing some consternation [124, 144-148]. Bradshaw [45] argued that what was required was the clear definition of national minimum measurable standards of competence for registration, which, Twinn and Davies [17] argued would benefit those charged with the responsibility of assessing levels of practice. As a consequence of such criticism, and based on the report of the UKCC Commission for Nursing and Midwifery Education[39], and research into the work of newly qualified staff nurses [149] an updated and clarified set of nursing competencies has been produced[11], under the main headings of

- Professional/ethical practice
- Care Delivery
- Care Management
- Personal/professional development

The UKCC have been anxious to ensure that these competencies are comprehensive, readily understandable, and applicable to all branches of nursing [115], but there has been no evidence of a systematic approach to their formulation. More importantly, the concern

over nursing and midwifery competencies has not simply been whether these are adequate and comprehensive. There has been widespread disquiet over the philosophical problems that have stemmed from having nursing curricula and assessments based on competencies.

3.8.7 Criticisms of competency based assessments

(a) Competency measurement fails to adequately assess nursing skills

Benner had misgivings regarding the appropriateness of competence testing in nursing, suggesting that core skills such as the ability to relate and empathise with others were precisely those skills most difficult to assess using a competence approach [150]. This concern has been shared within medical education [151]. However, it has been argued that attributes such as empathy which are not subject to direct observation, become detectable as behaviour and can be assessed through inference [59, 152]. This perhaps accounts for the absence of terms such as caring and empathy in the UKCC list of competencies to be achieved before registration [115]. These concepts have been central to the practice of nursing [153], and lack of their explicit mention in systems of competency based assessment is a major weakness.

In an analysis of competence approaches to learning in nursing, it was concluded that it was technically oriented and inappropriate to training people [124]. It was also argued that the distinctive character of an occupation becomes lost in competency based systems of training and assessment [122].

(b) Competency measurements rely on inference

The dependence upon inference in competency measurement has been criticised [122]. Inferences can be drawn in a number of different ways. The ability to transfer skills from one situation to another, the possession of affective attributes such as empathy, and in some systems the possession of propositional knowledge has been inferred from competence measurement, rather than measured directly [59, 122, 138, 152].

(c) Competence measurement is atomistic or reductionist

Alspach [58] argued that to be maximally effective, competency based education programmes need to emphasise the integration of the cognitive, affective, and psychomotor domains of clinical practice rather than focusing on the separate component elements of competence. This is very similar to Steinaker and Bell's theory of experiential learning (3.5.3). Competence measurements have tended to be reductionist or atomistic, through assessing the highly visible but often trivial, superficial and task-orientated aspects of an activity[122, 154]. This approach has typified the National Council for Vocational Qualifications system of assessment, widely criticised as being reductionist [155-157]. Conversely, with an extreme 'holistic' approach, there is the danger that relying on judgement alone precludes any analysis. Hager et al [122] concluded that in practice, some degree of atomism is acceptable, if it is within a holistic framework. Successful competency approaches that have achieved this balance have been holistic by acknowledging the complexity and demanding nature of professional activities (elements), and by taking account of the context of the activity. By integrating this approach with the attributes of the individual, competence comprises a relationship between the professional and their work [122]. Where such a balanced approach has merit, it may slip into the problems of an atomistic approach through the passage of time, or be de-valued through the differences in perception and approach of different assessors [155].

3.8.8 Conclusion

Nursing 'performance' referring to what an individual actually does in the practice context, assessable only through direct observation, is the most valid indicator of nursing ability. This distinguishes it from competence, capability, and competency, which can be measured in a number of context independent ways, but which have many drawbacks. The observational assessment of performance and how this process can be facilitated has therefore become a major focus of the multimedia program. Given the above discussion, an approach to the assessment of performance must be sufficiently atomistic in order to avoid complete reliance on clinical judgement and intuition alone, while being holistic enough to capture the true essence of a nursing performance. An integrated approach by Hager et al [122] solved this problem by identifying and assessing 'key elements' central

to the practice of nursing. An important role for Benner's [88] 'Domains of Nursing Practice' (table 2) has therefore emerged. This categorisation could be considered equivalent to Hager's 'key elements', referring as they do to aspects of performance that are fairly holistic yet capable of further subdivision. Existing between atomism and holism, and independently assessable, Benner's Domains of Nursing Practice were therefore incorporated into the instructional design of the multimedia program as 'key elements' of nursing performance to be demonstrated, modelled and assessed through simulation.

Chapter 4: The Observational Assessment of Clinical Performance

4.1 Summary

This chapter, in analysing the observational assessment of performance identifies useful parallels with the observational research method, providing an insight into issues affecting the reliability and validity of the tools and processes of observational assessment. How to construct and use instruments to aid the observational assessment of performance has been identified in the literature. This ranged from instruments developed by academics working in isolation, to much more rigorous approaches involving clinicians and reviewing the available literature. In the absence of a 'gold standard', the most valid approach has been to build an instrument on the basis of a theoretical framework representing the construct 'nursing'. Commonly, the works of Benner and Steinaker and Bell have been used for this purpose. Inter-rater reliability has also been an issue, but the use of a criterion referenced rating scale, such as that developed by Bondy, has been found beneficial.

The process of assessing practice has been devoid of an empirically established theoretical basis. However, one useful approach described by Wilkinson acknowledged the inter-related nature of assessment, teaching, and learning, in the educational context post-Project 2000.

The chapter concludes by identifying that interactive, varied approaches to teaching are required to prepare observational assessors of performance. One training model by Goldstein and Sorcher that considers role modelling, practice, social reinforcement, and transfer of learning was identified as having potential for guiding the instructional design of a multimedia program that could also include the works of Benner, Steinaker and Bell, Bondy, and Wilkinson.

4.2 Observational methods

Parallels between observational research, and the observational assessment of clinical performance can be drawn. For example, the assessment instrument in student assessment may be regarded as equivalent to the 'plan' that observational researchers use to guide data collection [2]. However, instrument construction, observer preparation and the general conduct of observational research has normally been with much more rigour.

Observation has been criticised as being subjective, time consuming, and of limited objectivity and empirical value [2, 158]. Yet one study by Johnson and Wilhite [159] compared student results in subjective clinical evaluations with objective tests and concluded that acceptable levels of validity and reliability can be achieved. Observer accuracy and objectivity can be improved by using structured observational methods, and tools such as rating scales and checklists [2].

It has been possible, theoretically, to establish the validity and reliability of clinical assessment tools, yet very few have been subject to this [48, 154, 160]. Little has changed in 25 years [161-163]. One identified problem has been the lack of a 'gold standard' against which to test new, or even established instruments [48]. Despite this, as curricula have developed and changed, new and better assessment instruments have been demanded by the NBS, in part to differentiate one curriculum from another [11].

4.3 Factors in instrument development

Establishing the validity and reliability of instruments used in the assessment of clinical performance has generally been neglected. Yuen et al [64], in 1987 found that knowledge of the effectiveness of different methods of clinical assessment was inadequate. A similar situation existed in the year 2000, when an NBS funded research project found that none of the assessment instruments then in use had been fully evaluated [48]. Before examining the psychometric properties of the available tools, the fundamental principles of classical test theory, namely reliability and validity, have been explored. By convention, an assessment is considered valid if it assesses what it requires to assess, and reliable if it measures with accuracy, consistency and precision[164, 165]. The inter-relationship of

these concepts has been acknowledged, as a measurement instrument lacks validity if it is unreliable [2].

4.4 Reliability

Measurement error is a major source of unreliability in research instrument design and use, and steps need to be taken to minimise its effect. It could be argued that factors that have been shown to contribute to measurement error, such as measurement instrument format and clarity, response sampling, situational contaminants and transitory personal factors[164] apply equally well to the assessment of clinical practice. Similarly, Guilford [166] cited six sources of error, each of which could apply to the observational assessment of performance (appendix 3). Examples that appear to have particular relevance to the criticisms made of student nurse assessments are errors of 'leniency' and 'central tendency'. The error of leniency, where the assessor is less critical than the student's performance warrants, has been claimed to lie in the assessor's attitudes. The error of central tendency, where the assessor may play safe and give an average rating, results in stereotypical, unreliable and invalid ratings of student performance. This can be a function of the rating scale design: an odd number of points have an easily identified midway point compared with those having an even number [73].

Reaching agreement between raters when assessing nursing skills has been problematic [73]. This improved when more easily observed behaviours were being evaluated compared to those that were covert, or complex. Bondy [167] similarly noted that raters were more lenient when judging behaviour that was abstract or open to subjective interpretation and for assessors to be more critical as observed behaviours became more structured and measurable. The insensitivity to 'low visibility skills', such as non-verbal communication, has been a weakness of observational assessment tools [48]. This is unfortunate, as these skills have been argued to be crucial to nursing (3.8.7a).

Establishing the reliability of an instrument has most commonly been through test-retest reliability, and tests of internal consistency[2, 168]. Achieving inter-observer or inter-rater reliability in observational assessment instruments, although desirable, has been

impractical because of the large numbers of mentors who may be involved. Although a record of the number of mentors in Scotland has not been maintained, the number can be estimated. As there were approximately 7000 indexed nursing students in Scotland in 1998 [13] and as 50% of an undergraduate course is spent on placement then the minimum number of mentors required would be 3,500. This conservative estimate reveals the difficulties involved in establishing inter-rater reliability.

4.5 Validity

In the context of the observational assessment of performance, validity has been argued to contain two distinct components; the intention of the assessor and the nature of what requires to be assessed [169]. There has been no simple way of establishing the validity of an assessment instrument. Different perspectives have therefore been explored.

4.5.1 Ecological validity

This has been concerned with how the results of a study apply to different contexts, recognising that different environments can produce different behaviours [170]. This would suggest that indirect assessment through reflective journals, and the assessment of competencies in a controlled environment, such as a simulation laboratory, do not have the same ecological validity as assessments carried out in genuine clinical environments.

4.5.2 Face validity

This has been described as a review of how good an item or group of items appear [168]. As work-based assessments have usually been grounded in the reality of the workplace, with their own rules, norms, and expectations then face validity has usually been claimed for any instrument produced within this context, to fit this context [7].

4.5.3 Content validity

This type of validity requires adequate sampling from the 'domain of content' [7]. This can cause problems if, as has been the case in nursing, the domain has not been clearly defined. An understanding of the boundaries of the content domain, and a strategy for sampling this has been required. Bennett [7] has argued that in the absence of a sampling

strategy, then adequate sampling may only be assumed if multiple sources of evidence have been pooled. This is consistent with advice given to those assessing students; through incorporating the views of other members of the nursing team, through asking hypothetical questions, and through the observation of practice, adequate sampling may be assumed. This approach forms the basis of the assessment of competence in many NVQ type courses[171].

4.5.4 Criterion validity

This comprises predictive and concurrent validity.

(a) Predictive validity

This is the ability of one measure to anticipate future performance, the best prediction of which has been argued to be prior performance [170, 172]. There have been no reported studies to determine predictive validity in the context of student nurse assessment. An instrument would have predictive validity if it can be shown to discriminate between those destined to fail and those who went on to succeed in future clinical assessments. Such information would be useful, but having identified a student 'likely to fail', the prediction would be thwarted, because of the ethical imperative to support the student. Assessment instruments based on Bandura's self-efficacy theory arguably have predictive validity, as measures of self-efficacy have been shown to correlate with performance [173]. However, there have been no reported studies of this theory being used in this way.

(b) Concurrent validity

This uses a pre-existing and accepted indicator, or 'gold standard' of the same concept with which measures can be compared. This has been lacking in student nurse assessment[48]. One approach would be to correlate students' clinical performance with that expected for the students' level of academic development but studies have found such direct correlation difficult to demonstrate. For example Gerrish, et al, examined forty-eight UK curricula post-Project 2000 and identified wide variation in the criteria used to define different levels of practice [66]. Although the existence of degree and diploma courses should indicate differences in expected academic attainment, in terms of

practice, outcomes have shown little difference [67, 174, 175]. This debate extends to whether diploma level practice is different, or should be different, to degree level practice. North American research has demonstrated a positive correlation between educational attainment and job effectiveness [129]; effectiveness in patient care delivery [176]; more advanced leadership, communication and nursing skills [177]. One piece of research by Bartlett et al indicated that in terms of leadership, diplomates scored higher than graduates on qualifying, but six months later graduates scored significantly higher on personal development, patient assessment, confidence and assertiveness [178], indicating that differences between diploma and degree levels may have a subtle effect not immediately apparent on registration.

Studies that analysed statements that purported to describe practice at degree, diploma and masters levels, and between novice and expert nurses found little distinction [67, 179]. Ashworth et al speculated that assessment statements could be interpreted differently at different levels, because language has not been sufficiently precise to unambiguously describe a level [67], nor is it sufficiently shared to reliably communicate a specification of level between nurse educators. The phenomena of nursing practice have been argued not to lend themselves to verbal categorisation. It would appear that attempting to establish the concurrent validity of a clinical assessment instrument would be pointless in this situation.

4.5.5 Construct validity

Construct validity tests the link between a measure and the theory underlying it [2]. The validity of the construct 'competence' could be examined by its relationship with other relevant constructs, such as 'knowledge', which underpins performance in the workplace [7]. A positive correlation through empirical studies would provide evidence of construct validity. The difficulty here lies in the different kinds of knowledge, 'practical' and 'propositional': knowing how and knowing that, the former resistant to empirical exploration, the latter more amenable. Some studies have found that cognitive domain predictors did not correlate with clinical performance [180, 181]. Rudolph et al [162], however, found a strong correlation between theoretical concepts and clinical

experiences. Construct validity, like concurrent validity has thus been difficult to establish, but isolated studies have demonstrated that it can be done.

4.5.6 Conclusion

The concepts reliability and validity are interrelated and are at the heart of research [164, 170]. The principles of carrying out observational research have been shown to have parallels in the observational assessment of student nurses' performance. However, the process has traditionally had much less rigour, making it difficult for conclusions to be drawn. There has been acceptance that no measurement tool is on its own is completely valid and reliable [161, 182], perhaps accounting for the dearth of published evidence of the reliability and validity of clinical assessment instruments. However, some instruments may be more reliable and valid than others, by the care taken in their construction. Some of the differences in approach are discussed next.

4.6 The process of developing an assessment instrument

4.6.1 Introduction

The tension in nursing research between measures that are reliable, and easy to apply, but which do not capture the complexity of nursing is evident [170]. Clinical assessment instruments that are simple, reliable, and easy to apply, for example check lists may not be valid because they do not assess the complexity of nursing. Conversely, instruments that adequately encapsulate and measure nursing ability may be so complex that inter-rater reliability diminishes, thereby reducing validity. Designing an instrument that is reliable and valid is never likely to be totally successful. The challenge faced in the instructional design of 'A Mentor's Guide to Assessing Student Nurses', has not been to produce an approach that is completely valid and reliable: such a search would be fruitless. The challenge has been to find an approach that achieves an optimum balance.

Several articles have described approaches to designing a clinical assessment instrument. For example, a four-stage process was described by Dunn [183] and Stecchi et al [184] outlined principles involved in developing their 'comprehensive evaluation tool'. The advice contained within these accounts, although by no means identical, indicated that the

tool needs to be flexible, to be usable in a variety of contexts, and to have developmental purposes. However, there was little acknowledgement of the need to establish reliability and validity, beyond content validity.

4.6.2 Conceptual frameworks: their role in the assessment process.

Given that the assessment of practice has to be based upon a sample of the student's overall performance [51], the assessment tool indicates those aspects of performance that need to be observed, and whether the 'units' of observation should be large or small. The assessment instrument therefore has a key role. For an assessment to be a valid representation of a student's nursing ability, and because assessment can drive learning, the assessment instrument ought to represent the construct 'nursing'. For this reason, some assessment instruments have been based on a nursing conceptual framework, in an attempt to provide construct validity. Within this approach, the assumption is made that important areas of performance are represented and sampled.

4.6.3 Conceptual frameworks used to guide the assessment of practice

The nursing process and other models of nursing have been commonly used to provide a structure for the assessment of practice [185]. Benner's Domains of Nursing Practice (table 2) has been popular. Gatley [99], for example, incorporated these within a district nurse curriculum, and Darbyshire et al [97] framed a nursing degree assessment tool around them. Gerrish et al [66] in a review of student nurse assessment found that either Benner or Steinaker and Bell (table 3) featured in the curriculum of 65% of the programmes analysed. In many cases the same conceptual approach extended to the assessment of practice.

Eclectic approaches have been taken by some; for example, Faltermeyer [110] utilised the work of Dreyfus and Dreyfus [186], Bloom[107], and Steinaker and Bell [106]within their assessment instrument. This approach has been criticised because of the lack of evidence to demonstrate that parameters within different frameworks intercorrelate [67]. The degree to which construct validity is provided is dependent upon the chosen conceptual framework being representative of the construct 'nursing'. Benner has undertaken

seminal work in this regard [90]. Selecting aspects of several frameworks could create a hybrid devoid of the philosophical underpinning of any model, and undermine construct validity. In this regard, and allowing for differences in philosophical perspective, Benner's theories have received wide acceptance (3.5.1). Creating an instrument that represents Benner's Domains of Nursing Practice could be assumed to have a degree of construct validity.

4.6.4 Instruments developed by faculty groups.

Many accounts of instrument development describe a process whereby a small number of lecturers or clinical members of faculty formed a working group, with the goal of producing a clinical evaluation tool for use in a particular situation or course [162, 187-193]. Reviewing the literature does not necessarily feature as part of this development process and pilot studies, if performed, ensured little other than acceptability and usability among the wider faculty and clinical staff. However, face validity could be assumed dependent upon the background and experience of the staff involved. Little else, in terms of reliability and validity was evident. Instruments produced through this type of approach may not be suitable out-with the local context.

4.6.5 Reviewing the literature and expert panel/ expert views.

This approach has typically begun by a group of experts (or a group in consultation with experts) reviewing the literature, before developing an assessment instrument. Some attempts at testing for validity and reliability have been reported in a number of accounts, but this has been limited to face and content validity [54, 194, 195]. Where clinicians were not represented on the expert panel during development, then face validity could not be assumed.

The process of deciding what requires to be assessed, and the key factors in the process of choosing or designing an assessment tool and strategy have been examined. It would appear that the most valid approach has been to adopt a theoretical framework that has wide appeal in nursing, such as Benner's. Details of a variety of assessment tools, as they have been reported in the literature, are described in the next section.

4.7 Assessment tool structure

4.7.1 The use of scales

Some approaches to assessment, for example, those that require the assessor to state whether competencies have been achieved or not can benefit from the use of a checklist. However, not all competency statements have been at the same level, making a checklist inappropriate. Girot [9], for example, indicated that UKCC competencies were much more holistic than NVQ competencies. Competencies that have been described as 'molar' cannot be assessed by a checklist, but require the use of a rating scale in order to increase the reliability in judgement [48, 167, 196]. Rating scales have also been used to indicate how much of a characteristic is present, and to describe how well it has been performed [138]. The labelling of such rating scales has varied considerably. In some, behavioural incidents have been used to distinguish one point on the scale from another [187, 197]. One such scale achieved satisfactory measures of reliability and content validity [198]. Norm-referenced labels such as 'good', 'average' 'poor' has been contentious in nursing, as these have resulted in the comparison of one student with others. However, criteria that directly relate to the performance of the activity have been shown to improve inter-rater reliability and this approach has been preferred in nursing [199, 200].

4.7.2 Rating scale structure

The acceptable number of reference points on a rating scale normally ranges from three to seven [138]. The greater the number of reference points, the finer the distinction that requires to be made when judging observations. This has been argued to reduce the reliability of the instrument particularly when only a sample of total performance can be assessed [199]. A range from three to seven points on a scale is less likely to contribute to poor reliability than poorly defined criteria [138]. In this regard, Squier [73] recommended thorough pre-testing to ensure that criterion statements are discriminable and ordered. A description of one such scale that has been tested with some rigour now follows.

4.7.3 Bondy's criterion referenced scale [196]

This has been described as a 'landmark' in the search for valid assessment criteria [48]. The 5-point criterion referenced scale, described in appendix 4, emerged from a review of the literature on student learning and skills acquisition, and discussions with clinicians. Faculty from four North American nursing programmes tested the accuracy and reliability of the rating scale by evaluating student performance on videotape [167]. Students were filmed performing a nursing activity at five different levels of competence: independent, supervised, assisted, marginal and dependent. The experimental group used the rating scale with criteria to evaluate performance; the control group used the same scale without criteria. The experimental group more accurately evaluated the students' performances. A brief description of several studies that demonstrate the value of Bondy's scale now follows.

Krichbaum et al [54] found that the validity and reliability of Bondy's rating scale contributed significantly to the ability of nurse educators to define students' clinical performance. Bondy's scale was integrated into an instrument used to assess 10 clinical learning descriptors. In a quasi-experimental trial, involving 85 subjects, this tool was judged to have internal consistency (Cronbach's Alpha = 0.97). A positive correlation with other outcomes of the course such as measures of moral reasoning provided some evidence of construct validity.

Hawley and Lee [201] studied Bondy's scale in the assessment of students at different points in their course (n=680). The authors found the tool to be objective, and the clarity of the criteria encouraged student self-evaluation. It was concluded that Bondy's scale provided the basis of a flexible tool for assessing students in a variety of clinical settings. This flexibility has allowed the scale to be adapted effectively, as Tower and Majewski [191] discovered when they modified and incorporated it into their instrument. Additional categories were developed from key themes in the curriculum. No psychometric analysis was included, however the authors expressed satisfaction that following two years of use, the tool was workable and discriminating. Similarly, Donoghue and Pelletier[202] adapted Bondy's scale. Specifying clinical contexts and stages of theoretical development

produced an instrument of greater complexity than the original. The appearance of measurement errors such as halo effects and lack of clarity in scoring criteria meant that the results of using the tool were mixed. However, the researchers found it a valuable starting point for further developmental work.

Norman et al [48] described several instruments derived from Bondy's scale[196], such as the King's Nursing Performance Scale [175], and a similar tool developed at Luton University[203]. Following a review of literature, Gilmore [154] acknowledged the impact that Bondy's scale has had as a method of increasing the reliability of assessors' judgement.

The use of Bondy's criterion reference rating scale therefore has validity that other approaches have not demonstrated. This scale has therefore been adapted for use in 'A Mentor's Guide to Assessing Student Nurses', and complements Steinaker and Bell's experiential taxonomy, as a means of describing different levels of skill.

4.8 The process of assessing clinical practice

4.8.1 Introduction

This section analyses how clinical assessments have been performed, with the aim of identifying an approach that is research based and most in keeping with the liberal educational philosophy of Project 2000. The key stages within the process of assessing students are considered.

4.8.2 Literature related to the process of assessing clinical practice

Analysis of the literature failed to uncover an approach to the process of assessing practice that is fully based on principles established through research. For example, Hepworth [63] described a model of the assessment process that focused on the part played by observation and perception. This model was claimed to be useful in describing 'judgement' as a simultaneous assessment of many different items of information or cues. However, it lacked any empirical basis, and although it concentrated heavily on the subjective nature of assessment, it did not adequately address the role of objectivity.

Similarly a 'process model' for the assessment of occupational therapy competencies seemed promising, but was overly complex and had not been empirically evaluated[59].

Underwood and Reid [204] described an approach that encompassed the teaching, learning, and assessing of both nursing theory and nursing practice, with frameworks for formative and summative assessment. The authors claimed this approach to be both reliable and valid, although no evidence of psychometric testing was provided. Taking a similarly 'inclusive' approach, Wilkinson, in a practical guide to assessing nursing students identified important elements [205]. This was based upon the author's experience in mentoring and assessing Project 2000 students. As such it has face and ecological validity. Details have been included in appendix (5).

Although widely regarded as reductionist, the NVQ approach to the assessment of competence [206] has outlined a process that need not be confined to competence assessment. Stages of evidence collection and judging have been identified.
Accordingly, evidence may be gathered from various sources such as

- Naturally occurring behaviour in the workplace
- Elicited behaviour
- Naturally occurring or elicited products from the workplace
- Simulations, competence tests, proficiency tests, projects.

Additional evidence may be required to supplement the above, by

- Oral or written responses to questioning
- Self reporting of prior achievement (through, for example, portfolio evidence) [206]

This approach seems comprehensive yet flexible enough to accommodate a number of different contexts. It is for this reason that this approach, together with that of Wilkinson [205], have been adapted and modelled within 'A Mentor's Guide to Assessing Student Nurses'. In addition to forming the basis of a model of the assessment process, and being used in the 'simulation' part of the program, the concepts 'evidence collecting' and 'judging' have been demonstrated separately because of their complexity.

4.9 Preparation for the observational assessment of performance

4.9.1 Introduction

The multimedia program, 'A Mentor's Guide to Assessing Student Nurses' (appendix 1) has been custom designed to be used by trainee mentors. However, the assessment of practice need not necessarily be performed by mentors. This section therefore focuses specifically on how assessors of nursing practice have been prepared, with a view to identifying successful elements that could be incorporated within the program.

4.9.2 Evaluation of courses of preparation

From the literature, it would seem that in practice, throughout recent history, qualified nurses have assessed learners with little knowledge of education or assessment theory. One survey of nurse instructors (n=280) found that many had been asked to perform their role with inadequate preparation [49]. This complements a much earlier RCN report that found ill-prepared students were being assessed by nurses equally ill prepared [207]. A similar situation was found in a larger, more recent study, by Neary et al [208]. The process of student assessment, the views of staff and students, and the preparation of assessors were analysed. Some assessors had no preparation and for others, the preparation was described as poor. The course lecturers were found to lack the knowledge necessary to answer assessor's questions, as has been found in other research [154, 209]. Neary et al also found that local policies to prevent inexperienced or inadequately prepared staff from assuming the role of assessor were frequently disregarded [208]. The results demonstrated that there had been no preparation for some assessors, with the majority suggesting that they had a little. Very few assessors believed that they had a lot of preparation. The overall impression from this research was that a systematic approach to developing skills in assessors of performance was lacking.

The ENB 998 course of preparation

This course was approved by the English National Board as preparation for those nurses, midwives and health visitors responsible for supervising, teaching and assessing students taking initial training for parts 1-6, or 8 of the professional register. It consisted of 15 days teaching which could be taken either together or as a series of study days throughout

a period of practice[210]. An evaluation by Crotty and Bignell [211] indicated that it was a valuable experience for participants. However, Neary et al [208] in their study, found general dissatisfaction in a range of preparatory courses, including ENB 998. There was reportedly too much theory, and too little focus on practical issues. More than half of the respondents considered ENB 998 to be a desirable preparation, but one quarter of assessors considered their course to be of no practical value to them, with the remainder wishing to see a new assessors' course. Interestingly, most assessors welcomed the idea of being assessed themselves by competent and clinically credible tutorial staff, and having feedback on their performance. This finding was similar to that of Duffy et al [14].

White et al [18] findings reflected widespread lack of confidence in ENB 998. However, this varied between institutions that offered it. It was recommended that a new national course be developed to replace ENB 998 to include the assessment of practice at different academic levels. This recommendation is similar to that of Watson and Harris [13], who concluded that, in Scotland, a Scotcats level three course was required. There has been little in the published research to indicate that a particular course or approach has all the elements necessary to prepare assessors for their role.

4.9.3 Methods of preparing observational assessors of performance

The literature search by Neary et al [208] revealed little evidence of recent research on how nurse assessors were prepared for their role. Erault and Cole [10] found in a large study of the assessment of professional competence, that training for assessors of work-based practice was generally provided. However, this was often inadequate, for example, lacking explicit assessment criteria.

Spool, in an attempt to identify the most successful training methods for observers of behaviour, began by reviewing the literature of the preceding 25 years [212]. A number of studies that compared different methods of training were cited to show that often no significant difference was found between methods, with regard to inter-rater reliability, halo effects and transfer of learning from the training situation to reality [213, 214]. Despite the available research being fraught with methodological deficiencies, such as the

absence of controls Spool concluded that training programs had the potential to be effective in improving observational assessment. In marked contrast to this, Newble et al [215] found that the only effective method of improving inter-rater reliability, and thus validity, was to identify unreliable assessors and remove them from that role. Training was found to be unnecessary for consistent assessors, and ineffective for those lacking consistency. It was implied that the personal characteristics of the assessor are important in how they perform their role, and that factors other than knowledge and skills, acquired in traditional ways, are important to the reliability of an individual's observation of behaviour. This will be discussed in relation to the attributes of mentors, and how these affect how this role is performed (5.5, p66).

A number of studies have employed liberal, student centred methods of training with some success. For example, Latham et al, [216], found that a training program whereby observers viewed video, practised, and discussed conclusions was more effective than simply receiving verbal instruction, or receiving no training at all. Other methods to maximise consistency among raters in nursing have utilised videotaped nurse-patient interactions as the focus of discussion and learning [167, 217, 218].

One approach that substantially reduced observational rating errors was described by Wexley et al [219]. Training workshops allowed assessors to practise by observing and rating interview applicants previously videotaped. Immediate feedback on the accuracy of ratings was given. Participant interest was maintained by using realistic situations and stimuli, and encouraging informal group discussions. This approach is consistent with a model described by Goldstein and Sorcher [220] (table 4), comprising four major activities argued to provide the conditions for effective learning.

Table 4: Goldstein and Sorcher's training model[220]

ACTIVITY	DESCRIPTION
Modelling	model persons portrayed (via videotape, normally)
Practice	trainees rehearse the effective behaviours demonstrated
Social reinforcement	provided in the form of constructive feedback
Transfer of Learning	enhanced by increasing the complexity and reality of learning situations.

This approach to training has attractive features that lend themselves to being used in a variety of ways, and with different media. It appears capable of addressing the weaknesses of recent practice in preparing assessors. It is also consistent with liberal, student centred approaches to learning, and by implication, the curricula of Project 2000. It has thus been adapted and incorporated into the instructional design of the multimedia program to prepare mentors to assess students.

4.9.4 Conclusion

From the literature discussed in this section, it has become apparent that observational assessors of performance have not always been given the preparation that the complexity of their role requires. Although much of the research is old and in many ways inadequate, more recent research indicates that little has changed. There is a strong indication that an effective course ought to involve different methods and approaches. One approach by Goldstein and Sorcher [220] incorporates features that have been found to be effective. Therefore, their training model has been used to inform the approach to preparing assessors in the program 'A Mentor's Guide to Assessing Student Nurses'.

Chapter 5: Mentorship

5.0 Summary

This chapter analyses the concept of mentorship, exploring its historical origins and its introduction to nurse education with Project 2000. The consequences of introducing a new set of roles and responsibilities for professional nurses have been explored. Findings in the literature identify semantic and conceptual variability. The classification of mentoring into 'true' and 'pseudo' mentoring types has been clarifying, as too have the models of mentoring proposed by Darling and Cameron-Jones and O'Hara. Conflict between mentor roles such as assessor and friend, and between mentoring and other professional responsibilities such as coping with a heavy patient workload, have become a disincentive to effective mentoring. Evidence from the literature suggests that mentor attitudes, knowledge, anxiety, and self-efficacy can influence how well mentors perform their role, particularly the assessment of students, the source of much conflict. Each of these has been examined from the perspective of measurement. Finally, courses of preparation for mentors, as described in the literature, have been reviewed. It was concluded that mentors have generally been poorly prepared, the duration and content of courses variable, with longer, more formal routes having greater merit than shorter courses.

5.1 Introduction

Mentoring has been an integral part of nurse education in the United Kingdom since the introduction of Project 2000 [221]. For example Gray and Smith [222] in their longitudinal study of Project 2000 students (n=17) concluded that the mentor role was crucial to professional socialisation. In many parts of the UK, mentors also assess students' clinical practice. The ease with which these roles have co-existed and how to model the relationship between mentor and mentee is worthy of exploration as 'A Mentor's Guide to Assessing Student Nurses' was developed for use within this context.

5.1.1 The development of mentoring in the nursing context

Mentorship is not a new concept. Mentor was a character from Homer's epic poem the Odyssey, who was appointed as teacher, guide, protector, advisor, and tutor to Telemachus, the son of Ulysses. This protective and encouraging relationship between a younger protégé and an older 'mentor' has been copied through the ages in academia, medicine, law, business and commerce [223]. It has much in common with apprenticeship, where skilled craftsmen have traditionally guided novices in the acquisition of skills [224].

Darling [225] has been credited with the widespread introduction and promotion of mentorship in nursing in the United States in the early 1980's. This was at a time when staff support systems in business were developed to promote the personal growth of employees [226]. This in turn followed developments in management theory, human resource initiatives, and the widespread acceptance of adult learning theories [227]. It was also considered a way of helping females to overcome business networks that constituted institutionalised sexism [228]. It retains a feminist orientation in some nursing publications [229]. For several reasons, therefore, when mentorship was adopted in the UK, a variety of interpretations of what it meant for nurses and nursing were made, preventing systematic research, or proper evaluation [116, 230] because variables such as length of placement, and the skills of mentors were difficult to control.

Mentorship has been reported positively in the literature, despite little evidence to show that having a mentor has any influence on a student's clinical learning [231, 232]. Indeed an early study of Project 2000 by Jowett et al [233] found that one third of students interviewed got little benefit from having a mentor. Role conflict and lack of time have been cited as major obstacles to effectively mentoring students, nevertheless mentors in studies claimed the relationship was mutually beneficial [231, 234, 235].

Mentoring has been viewed most commonly as an interpersonal relationship [236-238], distinguishing it from other roles such as preceptorship that have been described in more formal terms [239]. Yet a study by Walsh and Clements [224] found that only 40% (n=88) of experienced mentors either agreed or strongly agreed that they ought to function as a

student's friend within a close personal relationship. There appears to be a gulf between mentorship as conceptualised, and mentorship as practised. An exploration of some of the issues surrounding mentorship needs to be addressed, therefore, in order to understand this.

5.2 Defining the roles and functions of mentors: the '*definition quagmire*' (240)

When Project 2000 was introduced in Wales, for example, the nature and scope of mentoring roles and responsibilities was left to individual institutions (15). Similar devolved responsibility occurred throughout the UK. Consequently, the nature of mentorship has varied considerably (228, 241). Confusion and the absence of conformity concerning the concept of mentorship and the role and functions of the mentor have existed ever since (223, 227, 228, 239, 242, 243). Further confusion emerged through the use of terms such as preceptor, mentor, assessor, and supervisor to depict the same or similar roles (16, 231, 244, 245), when subtle differences existed (246). The appeal of mentorship, however, has not been diminished. Nor has it prevented the concept being adopted by other professional groups, such as teachers, dentists, the police force and fire service (227, 247, 248).

The semantic and conceptual variability surrounding mentorship has been held responsible for increasing the potential for role confusion, as new mentors struggled to make sense and differentiate between the large number of roles portrayed in nursing literature. Fowler (241) indicated that mentorship in nursing has varied from being a long term, inspiring apprenticeship, to the mentor being the person that accompanied the student on coffee breaks and showed them where to hang their coat. Butterworth (249) defined mentoring as

> '*an experienced professional nurturing and guiding the noviciate*' pg3,

This is similar conceptually to the ENB (250) who considered the mentor to be a counsellor and guide. Such loose definition has not been sufficient for those specifying the nature of mentorship or for those designing curricula and preparing individuals for

their role. Watson and Harris [13] and White [18], recommended the universal adoption of a more precise UKCC [251] definition of a mentor as

> *'a registered nurse, midwife, or health visitor who has the roles of providing support, guidance, and role modelling for students in the practice setting'* [251], p1.

It is worthy of note that the UKCC has not mandated that mentors assess students.

5.3 Mentoring in nursing, as practised

In most fields, mentoring has implied a dynamic, non-competitive and nurturing relationship lasting many years, encompassing choice, emotional ties and sponsorship [223, 252]. This has not been a useful concept in nurse education where placements can be as short as 4 weeks [13]. Barlow [243] described an early attempt to introduce mentorship into nurse education and concluded that mentorship theory was unworkable, as the relationship between mentor and mentee was not sustainable within the traditionally hierarchical structure of nursing in the UK. It is clear, therefore, that for mentorship to work within nursing undergraduate education it has had to be quite different to the ideal.

5.3.1 Types of mentoring

From a general state of confusion has emerged an awareness that there are different types of mentoring. Morton-Cooper and Palmer [227] coined these as 'true' and 'pseudo' mentoring. True mentoring has been further sub-divided into 'classical' mentoring, which is a natural, chosen, relationship with no defined programme the duration of which may be from two to fifteen years; and 'contract' mentoring, which is an artificial relationship created for a clear purpose determined by the organisation. The duration has typically been from one to two years, with limited options for choice of mentor. 'Pseudo-Mentoring' has marked differences compared with true mentoring. It implies mentoring in name only, with a focus on specific tasks of short-lived duration. It may involve guidance from several mentors for short periods, from six weeks to one year. The mentoring function as practised within undergraduate nurse education in the UK has most closely resembled 'pseudo' mentoring, although some placements have been less than six weeks.

5.3.2 Models of mentoring

These have encompassed the mentoring process and the relationship between mentor and mentee. A number have been created, but there has been no systematic analysis to determine if one is better than another [154, 232].

(a) The MMP: Measure of Mentoring Potential [225]

This was produced by Darling following interviews with 150 people, including 50 nurses, 20 doctors, and a number of healthcare executives [225]. Three requirements for a significant mentoring relationship emerged. These were attraction, action, and affect, within the context of fourteen different activities (appendix 6). Darling used these functions to develop an instrument that measured mentorship potential called 'Darling's MMP'[225]. This tool has been useful in identifying those with mentorship potential, but more importantly, it has demonstrated a capacity to identify the qualities and failings of individual mentors [253]. It has been difficult to use in the UK because of its size, and use of concepts and words that have proved to be culturally dissonant [248].

(b) Cameron-Jones and O'Hara's Model

Cameron-Jones and O'Hara[248, 254], adapted Darling's MMP together with the theory of Daloz [255] and developed an abbreviated version, following a study involving 87 nurse mentors and 39 student nurses. Eight key mentor functions were identified, half of which were broadly 'challenging' the remainder broadly 'supporting'.

Table 5: Cameron-Jones and O'Hara's model [255]

SUPPORTIVE ROLES	CHALLENGING ROLES
Friend	Model
Supporter	Assessor
Intermediary	Coach
Door-opener	Standard Prodder

A self-assessment tool based on these roles has enabled comparisons within and between professions[254]. It has also been shown to have predictive power, through identifying the learning effects of predominantly supportive or challenging styles of mentoring. For example, if the mentor demonstrates low levels of both support and challenge, then the student's learning would be static. If however the mentor provides high levels of both challenge and support, then the student's learning enters a 'growth' phase. High challenge and low support would cause students to retreat, whilst low challenge and high support would cause students to consolidate their learning.

The creation of a model that contains a balance of challenging and supporting functions probably indicates that Cameron-Jones and O'Hara [254] consider mentoring to be a combination of both. This conclusion is similar to that of Severinsson[256] who noted that supervision, leadership and mentoring necessarily contained an element of control. However, these analyses have competed with a number in nursing literature. For example, Spouse [257] in a naturalistic inquiry into effective mentoring, identified five predominantly supportive functions integral to the mentoring role. These were "befriending", "collaborating", "planning", "coaching" and "sense-making". This humanistic, supportive approach to mentoring is very similar to Darling's [225]. Fields[238] similarly identified major characteristics of mentors as role model, teacher, advisor, tutor, and provider of emotional support. Cameron Jones and O'Hara's model is therefore unique in its perspective on mentoring, balancing the tensions and conflicts that have accompanied mentorship in undergraduate nurse education.

The distinction between Cameron-Jones and O'Hara's conceptualisation and others, quite clearly parallels the distinction between pseudo and classical mentoring. This could explain why, for example, the "assessor" function has been included in Cameron-Jones and O'Hara's model [248] but not in the original MMP [225] from which it was derived. Cameron –Jones and O'Hara's model appears to be consistent with the pragmatism of pseudo mentoring in undergraduate nurse education, whereas Darling's approach has been consistent with the philosophy of classical mentoring. This has been apparent in the debate as to whether or not mentors should assess students.

5.4 Mentor and assessor: conflicting roles?

Whether or not mentors should assess students has been debated within the UK as practice varies from region to region. Arguably, where assessment has not been the mentor's role, as in classical mentoring, conflict can be minimised. However, when these roles have been combined, as it can be in pseudo-mentoring, then conflict has been shown to arise. Gray and Smith [222], for example, found students hesitant in asking for help from their mentor because they were afraid that it would bring their weaknesses to the fore, and would result in a wrong impression and perhaps a poor assessment. Anforth[258] similarly described the moral dilemma experienced by nurses as they assessed students at the same time as trying to be their counsellor, friend and guide. These examples have demonstrated the conflict between the supportive functions of the mentor, and their assessor role. It has been argued that the roles of assessor and mentor should be separate for this reason [242, 258] whereas others have argued that they should be combined[259, 260]. When the relationship between the mentor and the mentee is intense, personal, and emotional, as some have suggested it can be [261], then the assessor role sits uncomfortably with this and is at risk of being poorly performed. When the relationship between the mentor and the mentee is a more formal partnership as others have recommended [262], then the assessor role becomes a natural part of mentoring.

It has become clear that there have been mixed messages in relation to mentoring in Project 2000. The nature of under-graduate nurse education indicates a pseudo-mentoring approach to supporting students, an integral part of which is the assessor role. However, as has been the case in certain parts of the UK, by vesting the assessor role in an individual other than the student's mentor could indicate a wish to see classical mentoring relationships develop. This has caused confusion, and not surprisingly, students expecting to experience some form of intense, personal and supportive relationship with their mentor have been shocked to discover that mentorship was more about control than support [263].

5.5 Mentor attributes

5.5.1 Introduction

In practice, mentors' roles have been defined according to individual understanding, not necessarily based upon any of the original concepts [221, 264]. In addition, mentors' personal characteristics have been known to influence how well mentoring roles have been fulfilled[15, 257, 265]. As not everyone makes a suitable mentor, [266] the identification of 'mentoring ability' in individuals has been useful. Darling's MMP has been utilised in classical mentoring contexts for this purpose, and Cameron-Jones and O'Hara's model in pseudo-mentoring contexts. However, no instrument has been developed which has been shown capable of indicating how well mentors perform individual aspects of their role, such as the observational assessment of performance, which, as discussed in section 4.9.3, can be affected by factors other than knowledge and skill. Evidence from the literature now presented suggests that mentors' attitudes, knowledge, anxiety and self-efficacy can each affect how well they assess students, and may be useful as empirical indicators of assessment ability.

5.5.2 Mentors' attitudes

Attitude refers to a disposition towards or against a specified phenomenon, person or thing [267]. It has three components: affective, cognitive, and behavioural. As such, to have an attitude towards something implies a deliberate consideration of its value, a liking or disliking of it, and a tendency to behave towards it in a specific way [268]. Attitudes are gained through socialisation and experience, and are resistant to change. Mentors' attitudes towards students have been shown to affect the value of their relationship. For example, a phenomenological study of student nurses in community placements identified positive effects on learning when mentors showed interest in the student and their course [269]. However, not all mentors were interested in students. This may have been the result of pressures of time and conflicting responsibilities. How much this affected mentors was shown by Atkins & Williams [231] to depend upon whether mentoring was viewed as an integral part of their role as a professional nurse, and how much control individuals had over their workload. This study also found that mentors'

attitudes towards students and education varied widely, as did Twinn and Davies' [17] in-depth study of roles and relationships within Project 2000 courses.

Mentors' attitudes towards students were a major factor in Cahill's qualitative analysis of student nurses' experiences of mentorship [263]. Students believed that success or failure of mentoring depended on the mentor being friendly, approachable, interested in the student, and willing to give constructive criticism. When mentoring failed, reluctance on the part of the mentor to share knowledge and skills was typical. Cahill [263] concluded that when a relationship between mentor and students failed to develop this was more indicative of negative attitudes rather than through lack of time or motivation. Positive attitudes and mutual respect were considered essential for the relationship to function.

Wilson-Barnett et al [16] found several influential factors affecting clinical support for students. These were the mentor making time for students, being supportive, being positive about their course and being committed to teaching. Watson and Harris [13] found that mentors' attitudes had a crucial impact on student learning, and recommended that only those nurses who were motivated and valued learning should become mentors.

5.5.3 Measuring attitudes

Although there have been no studies focused precisely on this subject area, the available literature has indicated that mentors' attitudes, whether positive or negative, have a bearing on how their role is performed. For this reason, measuring a mentor's attitudes in relation to their role in assessment has potential as an indicator of ability and likely behaviour. The literature has identified possible items for such a scale, for example, attitudes towards failing students, and towards teaching.

Attitudes have been measured in many education and nursing contexts and examples abound [270, 271]. Oppenheim [272] provides useful guidance on attitude measurement and instrument construction using Likert scales. For example, an equal number of positively and negatively phrased items randomly distributed within an attitude measurement tool have been shown to help avoid errors of central tendency [272].

5.5.4 Mentors' anxiety

Anxiety has been described as a sense of threat to well being. This focuses on the perception rather than the reality of the threat that brings about anxiety. It is a subjective, emotional response to or cause of stress [267]. Stress, by contrast, is a biophysiological event. Anxiety is an essential state of arousal that serves as a motivating force in many different types of behaviour. However, even in otherwise healthy individuals, prolonged anxiety has been shown to result in elevated levels of adrenaline and cortisol, which have serious long-term negative effects capable of causing ill-health [273, 274]. High levels of anxiety can interfere with the performance of complex skills involving co-ordination and concentration [275]. There is evidence in the literature that mentoring can produce anxiety for a number of reasons. These are now explored in some detail.

(a) Anxiety through mentoring conflicting with other priorities

The extent to which clinical areas gain economic benefit from students for the contribution that they make to patient care has been estimated as £2 per student per hour for hospital placements [276]. A financial incentive therefore exists for students to be placed in wards. However, hidden costs in terms of increased workload have been borne by the staff who mentor these students [277]. Atkins and Williams [231] found that the extent to which mentors coped with additional and conflicting responsibilities as a result of mentoring, depended upon whether it was perceived as integral to their job, or as a separate additional responsibility. Those who accepted the role were less anxious about it, however, looking after students who were poorly prepared for their placement, for example, was a particular source of stress especially where the mentor had no control over other aspects of their workload.

Earnshaw [226] in a small study found stress to be problematic for newly qualified nurses, expected to function as mentors whilst still trying to establish themselves in a new job. Some staff nurses in this study felt intimidated by diploma students because of a lack of confidence in their ability to teach and supervise at diploma level.

(b) Anxiety caused through aspects of mentoring being incompatible

Being a mentor has been found to cause role conflict [223, 228, 230, 231, 234, 258, 278, 279]. In particular, the extent to which an assessing function conflicts with supportive functions, was noted in section 5.4. This situation has also been shown to be stressful for students [280]. A fundamental conflict arises through attempting to adhere to the incompatible philosophies of traditional mentorship and andragogy. At the same time as guiding, teaching, befriending, counselling and supporting students, mentors have been expected to encourage self-direction and autonomy [230, 281, 282].

(c) Anxiety caused by being responsible for students' assessments

Wood [49] surveyed 197 North American nurse educators and identified fourteen separate sources of concern, the most relevant being the legal implications of a contested assessment, and the difficulties caused by having to use unfamiliar assessment methods. Increased stress was found to be a consequence of having to assess students. Students have also been stressed through having their practice assessed[208, 283]. Anxiety in relation to clinical assessment has affected how the assessor has performed their role [205]. For example, Hayes [278] found mentors did not give students negative feedback because of their anxiety, especially in situations where conflict was anticipated.

(d) Anxiety through lack of preparation for mentoring

The importance of Project 2000 was discussed in the nursing media to such an extent that qualified staff felt nervous and unprepared for the mentoring role that was being given them [16]. This was worsened by not being given the information or skills with which to adequately perform the role. During study days and courses, lecturers confused rather than enlightened prospective mentors. Educational jargon on mentorship courses [13] and the lack of emphasis on practicalities created anxiety in prospective mentors who felt that their needs were not being met [154, 232, 284]. One community nurse described her initial experience of mentoring as "horrendous" because of her perceived inadequacy[285]. May and Domokos [28] similarly found that working without adequate preparation for assessing, evaluating and teaching was a major anxiety [286]. In addition, some mentors

felt ill equipped in their current practice through inadequacies in their own professional education [232].

(e) Mentors' anxiety in response to student stress

Marginality and stress affecting students have been recurring themes in Project 2000 research [30, 31, 287, 288]. New and stressful challenges associated with innovations in course design, frequent changes of placement, supernumerary status, and self-directed teaching methods have contributed [46]. Clinical assessments have been an obvious source of anxiety [263, 289], as too has the situation where the student has had poor relationships with their mentor [263]. However, a quantitative study that compared the stress experienced by nursing students pre- Project 2000 with students post-Project 2000 (n= 292) [290] demonstrated no statistically significant difference. It was interesting to note that students in both courses demonstrated the same pattern of increasing stress with time in training. Gray and Smith [222] found that for some students going on placement, this was an event tinged with both excitement and anxiety. This has been attributed to fear of the unknown [233, 291]. Stressful clinical environments have been a major factor for students to cope with, for example, one student on placement was reported by Wood [292] as anxious, lacking in self-confidence, and unable to cope with the complexities of patient care. The link between raised anxiety and reduced performance was observable in students on placement.

The relationship between anxiety and performance has been well established. Anxiety has major consequences for the performance of skills [273, 275, 293, 294], too much anxiety having been shown to be detrimental, as has too little. As the observational assessment of students is acknowledged to be a complex skilled activity [212, 231] it follows that excessive anxiety could adversely affect assessment performance. Conversely, if assessors of performance have not been sufficiently anxious, then the performance of a student's assessment is likely to be affected through apathy. Anxiety in relation to the assessment of performance therefore appears capable of affecting that assessment, either too little or too much being detrimental.

5.5.5 The measurement of anxiety

Anxiety has been measured in many contexts, ranging from patient anxiety in coronary care [295], to the effects of computer use on students [296]. Many different instruments have been developed, from the focussed Computer Anxiety and Learning Measure (CALM) [297], to the much more widely used State-Trait Anxiety Inventory (STAI) developed by Spielberger [298]. Streiner makes the point that when wishing to develop an instrument for application in an area that has already seen much scale development, such as anxiety, it is difficult to come up with anything completely new[299]. The STAI has been readily available on licence from Mindgarden ™ [300]. The advantage of this instrument is that psychometric testing has been performed, normative data readily available and its validity established [298]. As anxiety in relation to assessing students is 'state', rather than 'trait' then only one half of Spielberger's STAI would be useful in this particular context, the State Anxiety Inventory (SAI). This is a 20 item, 4point Likert-type scale, which is capable of being used either on its own or with other measurement instruments[298].

5.5.6 Knowledge

When assessing students, it would seem logical and appropriate that the assessor should know what to do, and how to do it. Knowledge of the curriculum has been found to be a key factor influencing the ability and willingness of mentors to respond to students' learning needs, to challenge students and to cope with being challenged in return [257]. It is therefore necessary for mentors to know about the clinical assessment tool, grading methods and the educational institution's evaluation strategy [49]. In addition, assessors need to be familiar with relevant learning objectives [205, 301] as, for example, Baillie [269] found that students benefited when supported by mentors who were knowledgeable about their course. Davis et al [302] noted that students expected their mentors to have this information. However, mentors' knowledge of these matters has been shown to vary considerably. Students have been supervised and assessed by mentors with inadequate knowledge of what the process involved [257, 303], which has resulted in some students complaining of unfairness [304]. Applying themselves to the task, and finding time to learn how to perform their role effectively is only one aspect of the problem for mentors.

For example, Bradley [284] found that some had difficulty in understanding how assessment instruments should be used and how the documentation should be completed. Unfamiliar, technical language and educational concepts have been thought to contribute to this [13, 208].

5.5.7 Knowledge measurement

Knowledge and understanding can be measured easily through multiple choice or short answer type questions [305]. Knowing about the assessment process and the instrument in use, are minimum requirements for mentors involved in assessing students. However, knowledge is not simply propositional in nature and cannot be fully determined through the recall of information. Determining 'know-how' in relation to assessment is an important indicator of ability. An indication of its presence can be ascertained by measuring the individual's self-efficacy in relation to the process of assessing students. The origins and usefulness of this concept are explored next.

5.6 Self-efficacy

5.6.1 Introduction

The assessment of clinical skills has been shown to be complex and multidimensional. It is not merely a cognitive process requiring knowledge and understanding, but is subject to the effects of psychosocial factors such as attitudes, and anxiety. These are major influences in their own right as discussed, but one theory; self-efficacy[173] appears to incorporate these factors. Central to Bandura's theory of self-efficacy is the belief that thoughts, beliefs and feelings affect behaviour[306]. It has been claimed that perceptions of efficacy influence the actions that individuals take, how much effort they invest, how long they persevere in the face of disappointing results and whether tasks are approached self-assuredly or not [307, 308].

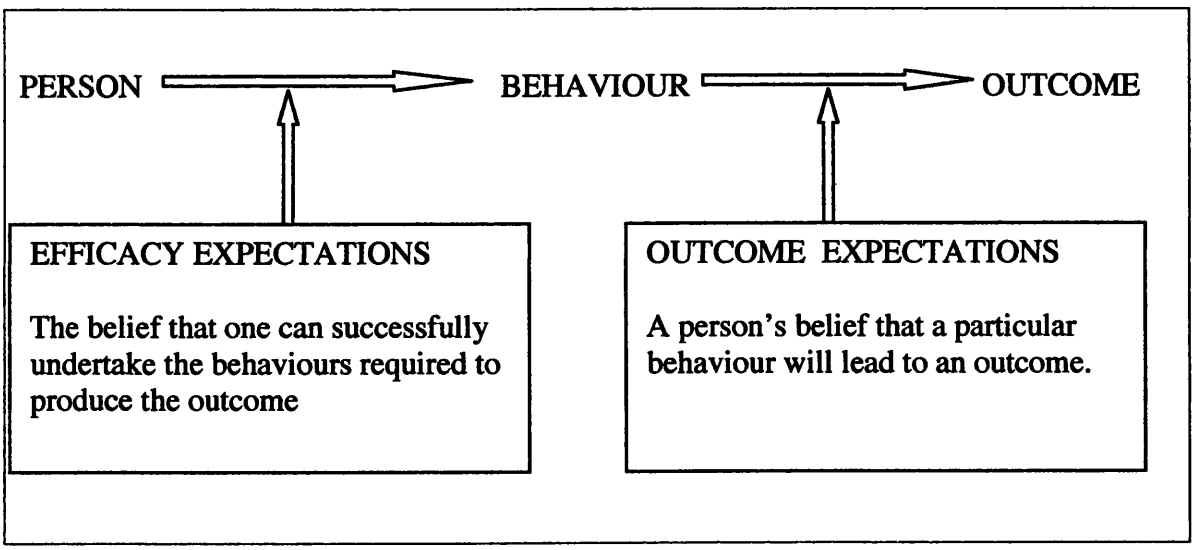
Studies have indicated that some mentors have been inadequately prepared and positive attitudes towards students and mentoring have been shown to be lacking, causing anxiety in those involved [13, 231, 257, 263, 269, 287, 303]. In this situation, it could be assumed that an individual mentor's concept of self-efficacy in supporting, assessing and teaching students has been low. Furthermore, it could be predicted that mentoring would not be

performed effectively. The concept, self-efficacy, is therefore worth exploring further, as it may provide an insight into the difficulties associated with mentoring and assessing student nurses. It could also indicate ways of overcoming these difficulties.

5.6.2 Origin of self-efficacy theory

Bandura [173] in seminal work proposed the theory of self-efficacy to help explain and predict human behaviour. He suggested that behaviour in most situations is mediated by beliefs about one's capacity to successfully perform a given task or behaviour. These he called 'efficacy expectations' to distinguish from 'outcome expectations' which are beliefs that the successful performance of a task will lead to a particular outcome.

Figure 1: Bandura's schematic representation of the distinction between efficacy and outcome expectations [173]



Self-efficacy theory has been tested in many disciplines and settings. In general, findings from research have supported the explanatory and predictive power of self-efficacy theory. For example, Hodge [309] found math self-efficacy to be one of a number of factors affecting nursing students' drug dosage calculations. Haycock et al [310] found that efficacy expectations were a significant predictor of procrastination in a study of university students (n=141). Zhang and Espinoza [311] demonstrated that attitudes towards

computers and computer self-efficacy were significant predictors of the need to learn computer skills.

Self-efficacy beliefs have been important influences on motivation and behaviour because they mediate the relationship between knowledge and action [308]. That is, cognitive, situational and affective factors influence behaviour partly by influencing self-beliefs and, as such, these beliefs are strong predictors of performance. Both efficacy expectations (the conviction that one is capable of actually carrying out the behaviour required) and outcome expectations (a person's estimate that a given behaviour will lead to a certain outcome) are incorporated in Bandura's theory of self-efficacy [173].

5.6.3 Self-efficacy and other 'self' theories.

Self-efficacy theory differs from other self-theories such as self-esteem and self-concept, although all have been concerned with global self-image. Both self-efficacy and self-esteem have been concerned with an individual's judgement, however the former deals with 'capability' and the latter with 'worth'. Self-concept is a collection of beliefs about oneself, which have not only influenced behaviour, but have also contributed to the understanding of how individuals feel about themselves[312]. Both self-efficacy and self-concept support the assertion that changes in 'self' lead to changes in behaviour, which contrasts with the view that changes in behaviour will lead to changes in 'self'[312].

5.6.4 Critique of self-efficacy theory

Some concern has been expressed regarding the distinction between efficacy and outcome expectations[313] (figure 1). The measurement of efficacy expectations can be confounded by outcome expectations, in poorly constructed measurement instruments[314]. Brady [315] cited the example of a commonly used arthritis self-efficacy scale [316], where some items intended to measure self-efficacy, actually measured outcome expectations. This is a major error as only 'efficacy expectations' have been found to correlate with performance [308, 317]. Brady [315] suggested that much of the published self-efficacy research had used poorly designed research instruments.

5.6.5 Self-efficacy experimental research: selected literature

Davis et al [318] used measures of knowledge and self-efficacy to evaluate a rheumatoid arthritis patient education program. The study utilised a self-efficacy scale that had been previously shown to have test re-test reliability, and acceptable internal consistency. Pre- and post-educational intervention measures were taken, with a group of 55 patients, followed by a third measurement after three months.

Knowledge and self-efficacy were shown to have increased significantly in an analysis of variance (ANOVA). There was no significant difference comparing post and delayed scores, suggesting that the improvement was sustained. The results need to be treated with caution as the chosen self-efficacy scale has since been discredited by Brady [315], and there was no control group, nor systematic approach to sampling. However, the research design was pragmatic, and applicable to the evaluation of any educational intervention, including multimedia.

Goldenberg et al [319] analysed the changes in self-efficacy of senior nursing students (n=36) and their preceptors (n=38), as a clinical placement unfolded. A 52 item scale was developed based on placement learning outcomes. As expected by the authors, the self-efficacy of preceptors was high initially and showed little change during the placement, whereas students showed a significant increase in self-efficacy as the placement unfolded. The authors claimed that the instrument was reliable, but offered no data to support this. In addition, there were no controls, nor random selection of subjects. However, as in the study by Davis et al [318], it was interesting to note that changes in self-efficacy were statistically significant immediately following the experience.

Murdock and Neafsey [307] used pre- and post-intervention measures of knowledge and self-efficacy, in an evaluation of a pharmacology course for qualified nurses. Data collection was repeated 2 years later. The statistically significant increase measured immediately after the course (ANOVA), was sustained at delayed post-test at 2 years. Reliability and validity were both claimed for the 96-item instrument, although details

were omitted from the article. Like the research on self-efficacy reported above, there was neither control group nor randomisation in sampling. Interestingly, like Goldenberg et al [319], the learning outcomes of the course were used as the basis for the development of the items within the questionnaire.

Harvey and McMurray [320] presented a much more rigorous and detailed piece of research than any so far described. Their work centred on the development of a research instrument to identify problems in nursing education and career progress in Australia. A pilot led to the refinement of the final instrument before being used with the main study, involving 306 students. Test re-test reliability, internal consistency, and construct validity, were satisfactory. It was found that low academic self-efficacy was predictive of withdrawal from a nursing course, whereas low nursing self-efficacy was not. This study demonstrated the rigour which is required to develop a valid and reliable self-efficacy scale, which, being required for each new context [321], is a major disincentive to the use of such questionnaires.

O'Halloran et al [322] in a small study of students (n=31) undertaking a research course measured self-efficacy in relation to nursing research, pre and post educational intervention. A five-point Likert scale was developed by subject experts, providing face and content validity. There was no evidence of test re-test reliability or measurement of internal consistency. Post-tests, like the studies by Davis et al [318], Goldenberg et al[319] and Murdock and Neafsey [307]were done immediately on completion of the course, and recorded a statistically significant increase in self-efficacy.

Schlessel et al [323] in a two-group trial (experimental(n=36) and control(n=47)) measured the knowledge, self-efficacy, and anxiety pre and post training in cardio-pulmonary resuscitation, in the parents of healthy infants/children. The experimental group received training, the control received nothing. Both groups were tested before and after the training program, and then again 1 month later. The instrument was a 7-point Likert scale for both anxiety and self-efficacy. It was subject to test-retest reliability, and found to be

within acceptable parameters. Results demonstrated significant increases in knowledge, self-efficacy, and a significant decrease in anxiety in the experimental group compared with the control group, immediately following the intervention. Delayed data collection after 1 month revealed no significant differences between experimental and control groups. When demographic factors (such as education, and previous experience of cardio-pulmonary resuscitation) were taken into account, the increase in knowledge immediately post-intervention was found not to be significant.

These studies, taken together, indicate that self-efficacy instruments require to be developed with rigour, and have acceptable validity, test re-test reliability and internal consistency. Pre- and immediate post- intervention study design has been shown capable of demonstrating statistically significant changes in self-efficacy, while a delayed measure yields little more of value. Self-efficacy research design could be improved by rigour in sampling, randomisation, using a large enough number of subjects, and including a control group.

There is some evidence that the concept of self-efficacy has been utilised in some evaluation research without the underpinning theory being recognised or acknowledged. Draper et al [324] described confidence logs as a useful diagnostic instrument in evaluating computer assisted learning (CAL) from a variety of perspectives. These confidence logs had the hallmarks of self-efficacy theory, although there was no reference to their theoretical basis. The instruments measured student confidence in having achieved learning objectives for a courseware package; on a five-point scale from *no confidence whatever*, to *very confident*. This was administered to students before and immediately after learning experiences, an increased score being interpreted as the student *believing* that they had learned from the activity. Draper et al [324] recognised that this was a very useful though indirect measure, which should be interpreted with some caution. Draper et al did not acknowledge the substantial body of research linking self-efficacy with performance. Perhaps if they had, they could have been more definite in their conclusions and less cautious in their recommendations.

5.6.6 Improving self-efficacy

It has become apparent from the preceding discussion that an individual's self-efficacy is not only measurable, but also amenable to change. The factors found to influence this are worthy of exploration. Perceptions of efficacy derive from past success or failure, observations of the successes and failures of others (vicarious experience), verbal persuasion, and sensations of comfort or discomfort in anticipating or actually performing a task [173, 307]. Therefore, providing a supportive learning environment where the individual can safely practice and be coached in order to develop positive experiences on which to draw, can be a successful strategy for improving self-efficacy. Mager [317] emphasised the importance of appropriate feedback to this process. It ought to focus on 'task', rather than 'self' to avoid a negative effect on self-efficacy that can be worsened by focusing on personal deficiencies. Role modelling has also been shown to improve self-efficacy. This effect can be enhanced if the model performs the same type of job, demonstrating what is required. If the model's sex, age, status, are similar to the trainee's then so much the better[317].

5.6.7 Self-efficacy and mentorship

Research on the effects of mentoring on students' self-efficacy has been limited. Myrick and Awrey [325] undertook a small study examining the effects of preceptorship on the clinical competence of baccalaureate student nurses. Twelve students were allocated to either experimental or control groups before a three week placement, during which only the experimental group was supported by preceptors. Pre-and post-tests of competence were compared, but no statistically significant difference was observed. Interestingly, the only significant improvement in the experimental over the control group was self-perception of performance. Preceptorship appeared to have improved self-confidence. This finding is similar to those of Wright [234], who concluded that the process of mentoring increased students' self-esteem and feelings of competence; and Marriott[116] who concluded that preceptorship improved self-confidence.

Duffy et al[14] reported a survey of the effectiveness of mentor preparation. Attendance at study days increased mentors' level of confidence in assessing the learner's level of

performance, completing assessment documentation, giving constructive feedback, teaching, and organising the students' learning experiences. The data collection instrument used was an adapted self-efficacy type Likert scale with three points: 'increased confidence', 'no change in confidence' and 'reduced confidence'. It is interesting to note from this study that half-study days were capable of improving the self-efficacy of participants.

5.6.8 Measurement of self-efficacy

Self-efficacy has been measured in many different contexts, from snake phobia to mathematics ability [321]. It has been shown to be predictive of academic performance[308] and a useful outcome measure of different interventions in health care [315, 326, 327]. Vispoel and Chen [321] reviewed 1,400 articles and identified 90 different self-efficacy measurement scales, 40% of which were education related and 20% health care related. These were designed for specific contexts and could not be used anywhere else without major adaptation, and the establishment of validity and reliability, limiting their usefulness [321].

Efficacy expectations have been shown to vary on the dimensions of strength, magnitude and generality [173]. 'Strength' has been described as how certain or confident the individual is of their efficacy [173]. Vispoel and Chen found that of 90 instruments reviewed, 62% used either a 'confidence' or a 'certainty' continuum for each item [321]. 'Magnitude' has been more difficult to measure. One method involved asking the individual to list a number of performance tasks in order of magnitude, indicating which tasks they considered themselves capable of performing [173]. This measure has been found to correlate positively with the strength of efficacy expectations. As 'strength' is an easier measurement, there seems little point in having separate scales for strength and magnitude.

The strength of self-efficacy determines whether a task is tackled, the amount of exertion invested, and the level of perseverance demonstrated in the case of setbacks [306]. The third dimension, 'Generality' refers to whether the efficacy expectation relates only to the

specific situation, or whether it can be generalised to other situations. Vispoel and Chen [321] found that this had been especially neglected, perhaps because of the difficulties in developing a tool sufficiently context specific to be accurate, while being applicable in more than one context.

Vispoel and Chen's literature review [321] identified a number of shortcomings in many existing measures of self-efficacy, including inappropriate scale formats, inadequate normative data, and limited evidence of test score reliability and validity. Reliability testing of published scales has generally been through measuring co-efficient alphas. However, the more rigorous test re-test reliability was rarely performed, neither was review by content experts, or pre-piloting and revision of items. Validity was generally assumed rather than systematically evaluated. This is similar to the findings of Brady[315]. Rigorous validation of any developed instrument was therefore indicated.

5.6.9 Conclusion

Despite the lack of standardised instruments and the need for rigour in the instrument development process, self-efficacy as a concept seems suited to the purpose of determining how well mentors assess students. An important error to avoid when developing a scale is to confuse outcome expectations with self-efficacy expectations. A second error to avoid is making the tool too generalised, as this has been shown to weaken its predictive nature [315]. Items to include in a data collection instrument could be identified from the literature review and adapted from the learning outcomes of mentor preparation courses. Mentors could then be asked to indicate how confident they were that they could perform the activity there and then. The 'confidence' continuum developed by Draper et al [324] seems simple yet appropriate.

In the context of the assessment of students' clinical performance, a useful self-efficacy instrument would distinguish between those mentors who were able to perform this well, from those who had difficulty with it. Self-efficacy measurement alone may provide an indication of actual performance, because of the factors such as emotional arousal and vicarious experience that influence it [173]. However, Schlessel et al. [323] demonstrated that triangulating self-efficacy with measurement of anxiety and knowledge provided

additional information. Their research evaluating the effectiveness of a cardio-pulmonary resuscitation training course, indicated that knowledge, anxiety and self-efficacy measurement sampled different domains. It would therefore seem logical that a robust instrument designed to indicate a mentor's ability to assess should include the measurement of all the influences upon it, that is, their anxiety, attitudes, knowledge, and self-efficacy.

5.7 Courses of preparation for mentors

Assessment of student nurses takes place within the wider context of mentorship, whether or not the roles of mentor and assessor are combined. It is worth considering, then, how mentors are prepared, and how much attention is paid to teaching skills in assessment within this overall context.

The adequate preparation of practitioners for a mentoring role has been shown to be vital[221, 231]. The reported benefits for students of good support in clinical areas through mentoring have been the development of self-confidence, improved learning, and reduced feelings of depersonalisation [140, 280, 328]. Yet, there has been no nationally recognised course to prepare practitioners for a mentoring role. Empirical evidence has highlighted the need for such a course. Neary et al [15] found that only 13% of practitioners involved in supporting students in Project 2000 courses (n=287) felt that they had been adequately prepared, and15% had no preparation whatsoever. For many, preparation for mentorship has been inadequate [17]. Other studies have noted the variable quality of courses, from area to area [16, 18]. There is ample evidence that the reality for many mentors is that they have had to learn on the job [13, 208, 232, 241, 329].

5.7.1.Content of mentorship courses

Neary [303] recommended that a new teaching, learning, assessing and evaluating course for assessors be developed. Watson [221] concurred, advocating a move towards prescriptive, realistic and consistent guidelines for mentoring to clarify the basic issues for all concerned. There have been numerous approaches to mentorship preparation, ranging from in-service half study days, to fully accredited Masters Degree modules.

Craddock [281] found that shorter, in-service programmes tended to focus narrowly on assessment, often limited to the correct completion of documentation. Fish[330] cited in Morton-Cooper and Palmer [227] indicated that during mentorship preparation, mentors ought to be provided with an awareness of rights and responsibilities, partnership agreements and obligations, knowledge of theory and practice, and the development of practice skills. Developing competence in the role of mentor is required, not merely becoming more knowledgeable about it.

It has been advised that mentorship courses should include curriculum information and details of the mentors' role in assessment [209, 235, 265]. Cerinus and Ferguson [331] gave a full account of course content for the preparation of preceptors (local term for mentor). Such things as course philosophy, principles of adult education, teaching and learning strategies, the learning environment and the understanding of the process of assessment, feedback, goal setting and evaluation ought to be included. In addition, Neary et al suggested the inclusion of theories of experiential learning [13]. These recommendations have been echoed in other research into mentors' needs [231, 332]. The ENB 998 course has been contentious as a possible preparation. Morle [228] considered this course appropriate only if teaching was regarded as the main responsibility of the mentor. Other criticisms of ENB 998, specifically as a preparation for the observational assessment of performance, have been discussed previously in section 4.9.2.

5.7.2. Duration of mentorship courses

Generally, practitioners undertaking longer, formal mentorship courses have felt better prepared than those who completed short specific courses [209]. Mentors who attended courses of 5 days or more were found to be significantly better prepared in all aspects of their role. This resulted in a recommendation by Watson and Harris [13] that mentorship preparation ought to be accredited at SD level 3. Brownrigg [333] found that a three-day mentorship course was too short to absorb the large quantity of information given to mentors. Some courses have been shorter than this. Laurent [334] described a two-day course of preparation, whereas Jowett et al [33] reported a range from single study days, to two-day workshops, five-day courses, and fifteen-day courses.

Duffy et al [14] presented evidence to suggest that experienced mentors require to be informed of curricular change and developments in the organisation of student education and that this has not been done well. The results of a survey into the effectiveness of the preparation and on-going support of mentors (n=150, response rate 47%) in one Health Board in Scotland indicated that mentors were reasonably happy with their initial preparation. However, for many mentors on-going support from both managers and academic staff was problematic. The study concluded that much remained to be done to provide mentors with the support required to meet their needs, although the nature of that support was unclear and warranted further investigation.

The problem of keeping practitioners abreast of educational developments has not been confined to nursing. For example, Williams and Webb [335] undertook research using the Delphi technique and found that practising radiographers were poorly prepared for their role in supervising students because they were unaware of changes to their course. Curricular change occurs regularly in nursing. As a result of the Peach Report [39], Project 2000 curricula have been radically re-structured. Changes of this sort have consequences for mentors, who ought to be informed of any such developments.

5.7.3 Conclusion

Mentorship in the context of UK nurse education should be acknowledged as a pseudo-mentorship model rather than the classical model. This would close the debate and confusion surrounding the role of the mentor, which has confounded its academic analysis and made the standardisation of mentor preparation almost impossible [263]. What has emerged from the literature is that mentoring is vital to the success of practice based learning post-Project 2000. Paradoxically, mentors have been poorly prepared to cope with a complex role that may be adversely affected by conflict, anxiety and negative attitudes. This situation could be enhanced by preparation that provides for growth in self-efficacy, positive attitudes and increased knowledge, thus reducing anxiety. Additionally, measures of self-efficacy, attitude, knowledge and anxiety have been shown to have validity as indicators of a mentor's ability to assess student nurses.

A method of preparation that provides consistency, depth and ensures quality whilst tackling those aspects likely to affect a successful assessment by mentors is required. One such approach, that may be capable of tackling this multidimensional problem in multidimensional ways, is multimedia. This is explored further in the next chapter.

Chapter 6: Learning with Computers

6.1 Summary

This chapter explores issues in relation to learning with computers, from pedagogical and pragmatic perspectives. Evaluative studies suggest that computers have the potential to improve the effectiveness and efficiency of learning, while accommodating different learning styles. Advantages in preparing mentors to assess have been identified, such as providing consistency in educational approach, increasing access to information for those unable to attend a formal courses and presenting an educational approach philosophically consistent with established curricula. However, learning advantages have not been demonstrated in every study: many comparative studies have been flawed, and many have shown no significant difference between computer assisted and traditional educational methods. The evolving use of learning technology in nurse education, the educational properties of multimedia, and instructional design considerations have therefore been critically examined. The crucial role of motivation and simulation has been identified, as too have user attitudes and anxiety.

The value of an eclectic educational approach to instructional design has been discussed, acknowledging the usefulness of different pedagogical approaches at different times and for different students, as outlined in Jonassen's taxonomy of levels of learning linked to instructional design.

Issues in relation to the evaluation of learning effectiveness have been explored, with difficulties identified. The use of the Randomised Controlled Trial in summative evaluation, the 'gold standard' research method has been examined, with advantages and disadvantages identified.

Finally, the usefulness of strategies for formative evaluation throughout the production phase of the program have been discussed, particularly Tessmer's stages of formative evaluation and categories of evaluation data.

6.2 Terminology

Computer Assisted Learning (CAL) has been synonymous with Computer Based Learning (CBL), Computer Based Training (CBT), Computer Assisted Instruction (CAI), and Computer Based Instruction (CBI). Most early CAL programs were simple text based tutoring systems, and linear in structure, although a small number included sounds, animation and graphics. Rudimentary simulations and games encouraged the exploration of concepts, but this was limited to drill and practice, multiple choice questioning, and exploration of dialogue [24]. The term 'multimedia' was coined in the early 1990's and applied to hypertext programs that were interactive, non-linear in structure and utilised a variety of presentation formats, including video, still photographs, animation, text, graphics and audio [336].

6.3 Modes of delivery of computer assisted learning

It has only recently been possible to utilise the full potential of different media in multimedia programs, by overcoming limitations such as portability of software, system memory and processor speed. The CD-ROM (Compact Disc, Read Only Memory) became available in the mid 1990's and became a popular method of copying and distributing software because of its large memory capacity, ease of access to data, and considerable installed user base. This was an important development, because complex multimedia programs with large file sizes could then be packaged and distributed cheaply and easily. Early technical problems such as digitising analogue signals, processor speed, and lip synchrony have been overcome. A newer, more sophisticated CD-ROM, the DVD (Digitally Versatile Disc) has much greater potential than the original CD-ROM, because it has much greater memory capacity. Similarly, the use of the Internet, especially using high-speed broadband networks, has the potential for delivering high quality, interactive learning packages to the user. This has advantages over previous methods, in that the content can be within a dynamic system, and easily updated. The availability of large amounts of data, in the hands of individual learners, has given interactive multimedia its potential to enhance learning[19].

Within this dynamic digital environment, it was increasingly feasible to produce an educational multimedia program that could be used by mentors in the workplace. It is, however, one issue making a program available and usable, but ensuring that learning occurs can be problematic [337]. Conflicting views and interpretations of research findings have raised issues in relation to how CAL programmes ought to be used in nurse education. This will now be explored.

6.4 Computers and nurse education.

Computer programs designed to assist learning have been available since 1966, but CAL programmes only became widely used in nurse education in the early 1980's, and were promoted through initiatives such as the Nightingale and ENB CAL projects [338, 339]. Much was learned from the experience of specifying, designing, producing and introducing computer assisted learning programmes. However, much of this was painful, as the product of much work did not live up to expectations. As Rowntree [340] stated in 1986:

"Computer assisted learning is the medium of the future...and it always will be".

It became apparent that students were much more sophisticated learners than the first wave of CAL programs and learning technologies allowed them to be. Learning technology has since evolved along two main divisions. The first has been through adapting hardware and software, initially developed for use in the wider world of business, to teaching and learning situations. Examples are spreadsheets, databases and the Internet. The second is courseware developed for specific instructional purposes. Use of the former category has grown rapidly in the drive towards globalisation of higher education whereas the educational role of the latter has been more modest [341]. This has been considered the result of a combination of factors. Information and Communication Technology (ICT) has increased in power, flexibility and compatibility, and costs have come down [19]. This has spawned a new generation of web-based educational applications to facilitate academic discourse, such as First Class™, WebCT™ and Blackboard™. In an education policy report to the UKCC, McNay and Bailley cautioned

against providing distance learning courses mediated through ICT without offering an option to students. They pointed to evidence of gender bias towards men, whereas many women (and by implication, many nurses) prefer the social aspects of traditional teaching and face-to-face contact [23]. Atkins [342] too expressed caution, arguing that a computer could not capture the intuitive and creative aspects of thinking and encompass all the values, attitudes and social aspects of learning. Schwirian [172] acknowledged the difficulty of defining and being explicit about values such as 'knowing' and 'caring', identifying a dilemma for developers of software that requires to conceptualise and model constructs at the heart of nursing.

This argument has been strengthened by critical commentary on ICT facilitated learning across the Internet. Many applications have been unsupportive of anything other than browsing a static set of links not customised to the learner's particular needs in any way [343]. Laurillard described such browsing as non-interactive activity [344]. However, computers have the capacity not simply to facilitate inter-personal communication, or as a method of sorting, storing, and providing access to information, but they can help develop metacognitive skills if appropriate pedagogical principles are used in their construction[345]. It is these aspects of learning technology that appear not to have fulfilled their potential, as the focus has moved towards computer-mediated communication.

6.5 Curricular issues

The uptake of communication and information technologies in education has been disappointing [346, 347], despite their availability and potential economic and educational benefits well understood. One explanation has been the culturally bound 'not-invented-here' syndrome [24] that has precluded some institutions from utilising the products of another. Another possible explanation has been that incorporating programs into existing curricula can be problematic. The difficulties in doing so have sometimes given the impression that CAL is supplementary, or for remediation [348]. This has affected the dissemination of high quality, and expensive, learning resources, such as those produced in the TLTP initiative [349, 350]. Variations between institutions and the absence of a

common curricular framework have been fundamental problems, resulting in applications designed for widespread use lacking curricular 'neatness of fit' [349]. Herein lies an economic catch: to produce a context-specific program that has high utility value can be very expensive, with little opportunity to recoup costs. Attempting to produce a program that appeals to a large market loses context specificity and the utility value drops, reducing the appeal of the program to those who may otherwise have found it useful. The reasons that prevented mass utilisation of CAL in the 1980's, such as a small potential market, poorly developed user skills, poor ease of fit and the not-invented-here syndrome have persisted [24, 351].

The overall design of the curriculum is recognised as being important to effective learning[352] [342]. This has implied achieving a balance and using complementary teaching methods, including CAL. However James and Turner[351] claimed that more emphasis, and investment, has been placed in hardware and software than in the human resources needed to capitalise on their potential. These elements, essential to any strategy, have only recently been acknowledged in the literature [353-356]. It is not surprising that when schoolteachers were asked about the impact of multimedia in the education of children, it was found that it contributed very little to the development of information technology skills, and in many cases contributed nothing to learning [357]. It might be reasonably asked, therefore, why invest and develop computers in education? Three possible reasons for computer use in education have been identified [358]. Firstly, it gives teachers a positive image, secondly, it produces working efficiency, and thirdly, it makes otherwise difficult activities possible, such as collaborative writing or the creation of simulations. There are clearly organisational, pedagogical and management of change issues to be considered when developing and introducing CAL, in order that potential benefits may be realised. Pedagogical issues will be returned to, but firstly the way in which the attributes of users impact on the uptake and utilisation of technology will be considered.

6.6 The effects of computers on learners: attitudes and anxieties

6.6.1 Attitudes towards using computers

The use of computer technology has been known to be affected by the attitudes of users [359, 360]. Change in users' attitudes has therefore been a readily available, quantifiable means of evaluating software. This measure has been used widely [339]: an improved attitude following exposure indicating an effective product, and vice-versa. However, Kay [361] argued that many computer attitude scales have questionable validity, through lacking any theoretical justification for the constructs used. One study, however, found remarkable similarity when a number of tools were tested for reliability and construct validity [362].

Some studies found that attitudes towards computers became positive after using them. Schwirrian et al [363], in a study of attitudes to computer use among nursing students (n=353) and staff nurses (n=358), found that the more computer experience individuals gained, the more positive their attitudes became. This finding is supported by Schare et al[364] who, in a controlled trial, found that students who learned through interactive video had a more positive attitude towards learning than those using traditional methods. Hamby [365] in a similar study found similar results relating to attitudes towards CAL: the experimental group, featuring computer use, had a statistically significant improvement in attitude towards CAL as a teaching method.

Lowdermilk and Fishel [366] in a Randomised Controlled Trial to evaluate computer simulations as a measure of nursing students' decision making skills, also used a 36 item attitude scale to evaluate students' attitudes towards computer assisted instruction as a learning activity. Pre and post-test comparisons for both experimental and control groups found that more than two thirds of respondents (n=70) had positive attitudes towards the use of computerised simulations.

Jensh [367] reported on a study at Jefferson Medical College where 223 medical students used CAI and IVD (interactive videodisc) tutorials in histology. Students developed highly positive attitudes towards the course and its content. Similarly, Soled et al [368]

compared the attitudes of nursing students towards learning, by comparing the attitudinal responses of student nurses who had received either a lecture or instruction through interactive videodisc on the subject of diabetes. Those students who received interactive videodisc instruction had significantly more positive attitudes than those who had not. Saranto et al [369] in a survey of 600 Finnish nursing students (response rate=62%; n=373) found an association between the number of teaching sessions using computers and positive attitudes towards the educational use of computers. Attitudes were found to be very positive on the use of computers in health care. This finding is supported by VanDover and Boblin [370] who similarly observed that students were more interested in clinical as opposed to academic applications.

Some studies, in contrast, have indicated negative attitudes towards CAL. For example, Day and Payne [371] found that student nurses being taught health assessment actually preferred traditional teaching methods, as they found this less frustrating and more enjoyable. Thede et al [372] noted that student nurses were reluctant to use CAL, with only a 10% spontaneous utilisation before it became mandatory. Baldwin et al [373] similarly found that students were dissatisfied with CAL and preferred teacher demonstration of psychomotor skills when learning nursing skills. Of interest was the finding that 26% of these students considered CAL inappropriate for any classroom lesson. A substantial number of respondents in Saranto et al's [369] survey (43%) thought that most nurses' attitudes towards computers were negative.

On balance there seems to have been more reports of attitudes becoming positive on exposure to CAL than vice-versa. This is an interesting finding, but it may not be a useful indicator of the quality of any particular program, or of outcomes in terms of learning that has occurred. It is confounded by the norms, expectations and values of the user group; for example, it would be difficult to imagine that additional computer use, no matter how good the program, would always result in improved attitudes. The measurement of student attitudes towards computers generally, or 'A Mentor's Guide to Assessing Student Nurses' as a computer program, could not therefore be used as a reliable indicator of the program's effectiveness. The interpretation of attitudes towards

computers and software, like opinions of their usefulness ought to be treated with caution (339, 374).

6.6.2 Computer fear, anxiety, and phobia

Computer anxiety is a complex phenomenon associated with computer attitudes. Both have been shown capable of determining an individual's behaviour and feelings towards computers (375). Following the poor public response to the mass availability of computerised technology in the late 1980's, instruments designed to measure computer anxiety proliferated (297). Surveys demonstrated that approximately 30% of the business community were anxious about computers (376). The term computerphobia was coined to explain why individuals were afraid and resisted using computers, having been given the opportunity to use them. This fear has been estimated to affect between one quarter and one third of the general population (375). Not all fears are phobic, or irrational, however. Computers can create the perception of threat, such as perceived loss of control. Such fears have been described as computer anxiety, the characteristics of which are outlined in table 6.

Table 6: Characteristics of computer anxiety: Brosnan and Davidson(375).

Avoidance of computers and computer areas
Excessive caution with computers
Negative remarks about computers
Attempts to cut short the necessary use of computers

This anxiety has been described as a form of state anxiety (298). Exploring the correlates of computer anxiety, such as gender, age, and experience with computers, shows that it is more prevalent in women than men, especially those choosing arts courses rather than sciences, and the incidence increases with age (375). Some studies have reported widespread fears and frustration regarding computer use among student nurses. This is not surprising given that nurses are predominantly female, are more likely to have studied the arts than sciences, and the proportion of mature entrants is increasing (23).

One real threat in teaching and learning situations is that differences between individuals using computers become apparent and exaggerated. Staggers and Mills [377] noted that data retrieval performance between individuals was as high as 20:1 using computers. In contrast, performance difference for non-computer tasks was no more than 2:1 for 95% of the population. Similarly, the quality of the outcome or product of computer use can demonstrate large differences between individuals. These factors may be significant in the context within which computers are used in education, and could be an argument for students working in isolation, rather than within groups. Saranto et al [369], for example, found in a survey of 373 nursing students, that 50% wished to study on their own, and 43% wished to pair with a colleague. This contrasts with studies where some students found working in pairs to be an effective coping mechanism to minimise anxiety [378], and given the choice, the majority of students prefer to work in a small group [379, 380]. This tendency has been noted particularly in female students [23]. Some students found groupwork utilising CAL programs an enjoyable experience [381, 382]. It would seem reasonable that given these conflicting findings, that students ought to be given the choice of working alone or in groups. This would allow students to utilise individual coping mechanisms to minimise any perceived threat that computers have been known to pose [382].

The assumption has been that succeeding generations would feel more comfortable with the use of computers and technology in education [23]. However, technology is finding new applications in education, for example, video conferencing, virtual reality, and high fidelity simulations, that are quite different to home and entertainment applications, and may always be new and threatening for successive generations of students. It needs to be considered whether or not computer assisted learning ought to be encouraged, made mandatory or optional. In order for this to be resolved, there now follows a discussion of the learning effects that have been found using this technology

6.7 Evaluating the effectiveness of learning with computers

6.7.1 Reported effects of using computers to learn

Miller conducted a review of 30 studies involving corporate, institutional and government users of interactive computer technology and identified learning advantages, for example, reduced learning time by up to 50%, instructional consistency, and increased retention of learning [383]. However, methodological difficulties have marred much research. Frequently, no significant difference in measurable educational outcomes have been found using computers, compared with traditional teaching methods [19, 356, 366, 384-388]. Typically, studies using small sample sizes yielded results that did not reach statistical significance [339], however, several meta-analyses of small studies have been useful. One, involving 199 comparative studies investigating the effectiveness of computer-based instruction in schools and colleges [389] resulted in conclusions similar to those of Miller [383], identifying mainly positive learning effects.

In nurse education, two meta-analyses, one by Cohen and Dacanay [390] using 29 studies, another by Belfry and Winne [391] incorporating 11 different studies, reached similar conclusions. Students, who used CAL, learned more and in shorter time compared to those students who did not. Knowledge retention was similar to traditional teaching methods, but students using CAL had statistically higher scores on tests measuring application of knowledge compared to those that received only lectures. These findings have been mirrored elsewhere in the health care education sector; for example, Pedlar[392]compared a CAL program with traditional lectures as a means of teaching oral surgery, and found a statistically significant increase in knowledge. It has also been demonstrated that students increased their problem-solving abilities using CAL [393]. Learning benefits have not been confined to the cognitive domain: some studies have demonstrated benefits in affective [394] and psychomotor [395] domains.

Not all analysts have viewed the results of the available research as positively. Schwirian [172] summarised findings by stating that CAL is an expensive activity that is neither superior nor inferior in terms of learner acquisition of information and skills. However,

less learning time may be required, and some like it, some do not. Clark [396] pointed to the lack of evidence of learning benefits of one learning medium rather than another, attributing this to methodological difficulties and the confounding effects of variables that were not controlled in small studies. Lewis et al found it difficult to derive substantive evidence from their review of published CBL studies [339]. Difficulties were categorised as intra-study variability, inter-study variability, and the trend of technological advancement. Within the reviewed literature there were few examples of experimental research into the effectiveness of CAL which combined control groups, randomisation and a large enough sample size.

Mayes[397] has taken a quite different perspective. He claimed that most benefit from CAL developments come to the instructional designer who gains an intimate knowledge and understanding of subject content through the process of construction. Learners who use the material as coursework do not learn as much. Taking this to its logical conclusion, students ought then to be involved in structuring their own learning packages and courseware. These more critical positions have superseded earlier, much more positive views on the benefits of CAL, such as those proposed by Procter [348], which are summarised in table (7).

Table 7: Perceived benefits of computer assisted learning: Procter [348],

For Students	having the opportunity to be self-paced, and to repeat parts of the program as desired the generation of self-study principles the identification of additional learning resources the challenge to develop knowledge and skills in a given area
For Teachers	the ability to identify the path a student takes through a program, allowing for personalised tuition a better awareness of student learning an ability to set a standard in a particular topic area.

CAL has often been considered a useful alternative to, rather than a direct substitute for traditional teaching methods, especially when no significant difference between methods has been found [355, 385]. It therefore seems possible, in some circumstances and for some users, to gain educational benefits through computers. Trials of sufficient size are therefore necessary in order to control the variables that can obscure and confound any findings. This and other methodological issues are discussed next.

6.7.2 Methodological issues

The evaluation of CAL has been contentious, because of difficulties undertaking comparative studies. A review of 26 evaluative studies spanning 30 years found significant design flaws, including small sample size, and lack of controls [339]. Using the rationale that CAL is one factor in a complex situation, it has been argued that a multidimensional and flexible approach needs to be taken to its evaluation. Such an approach would attempt to show how CAL contributes to the overall learning experience, and has been widely accepted [324, 349, 350, 398-400]. One flaw in the argument for taking this approach is that a carefully constructed, pedagogically sound program can influence learning, irrespective of context. Focusing primarily on the overall context diminishes the significance of the impact that the instructional design of the software can have, and does not indicate which parts of the software are more effective, and for what reasons. For example, elements such as navigation, provision of feedback, use of advanced planners have been shown to be of crucial importance to the learning effectiveness of programs [342]. Such information has been made available through the laboratory approach of studies into human-computer-interaction, (HCI) characterised by controlled conditions, and random assignment of subjects to experimental and control groups [401-403].

The point frequently made is that it has been difficult to remove confounding influences on scientific investigation into the use of teaching materials in real world settings, such as instructional method, and novelty. The number of evaluative studies with such methodological flaws lends weight to this argument [339, 396]. However, it ought to be possible to design rigorous scientific experimental research using randomisation and controls, in a pragmatic way. The randomised controlled trial as the 'gold standard'

research method [404, 405] has thus been chosen to evaluate the educational effectiveness of 'A Mentor's Guide to Assessing Student Nurses'. However, the program must be sufficiently well developed using appropriate learning theory to permit a trial without ethical constraint [406]. The instructional design of the program must therefore be theoretically sound, and the quality of production ensured by the implementation of a rigorous strategy of formative evaluation. How these issues have been addressed in the development of computer programs is discussed in detail, next.

6.8 Producing multimedia: learning theory and instructional design issues

6.8.1 Introduction

The capacity to respond to individual learning styles has been considered the greatest potential of interactive media [19, 407]. In addition, multi-sensory input has been known to enhance learning, and teaching has been argued to be more effective and motivating if a variety of media and techniques are used [305]. This has led to some program developments being driven by the ability to provide innovative alternatives in information technology rather than being based on learning principles [408]. This may be because of the abstract and complex nature of many learning models [409]. The importance of motivation, interaction and simulation in the development of sound pedagogy in the construction of a multimedia program capable of achieving desirable learning effects are now discussed.

6.8.2 Motivation

Multimedia has been described as exciting and engaging, through an array of sensory input. It can vividly and effectively simulate processes and events from the past and present [381]. These characteristics may be important in capturing the attention of the student and in providing motivation in simulated environments [382]. This is important, because if user motivation has not been addressed in the design and development of software, then learning materials, no matter how good, may be ignored [409]. Of themselves, however, motivational features are insufficient to ensure sustained engagement and retained learning [337]. In addition, it is not necessarily true that CAL works best when it is exciting and stimulating. Sometimes students learn more from

answering text-based questions than they can from interacting with a 'multimedia extravaganza' [381].

6.8.3 Simulation

Computers have been shown capable of providing an exceptional simulated environment by allowing the manipulation of many variables within a problem [410]. Even if definitive answers to a problem cannot be found by the user, a better understanding of the issues can often be gained. The use of simulation in education dates back to the early 1900's when it was used widely in military training [411]. In nursing, computer simulation has been advocated as a non-threatening and safe situation for learners to develop critical thinking skills [87, 412]. However, Lowdermilk and Fishel [366] found no difference in the clinical decision making skills of those students undertaking CAL simulations in a randomised controlled trial (n=64), compared to those who did not. In engineering, Edward [413] found that although a computer simulation of an engineering plant was considered an acceptable alternative to conventional laboratory work, much was lost, such as practical appreciation and the development of team-working skills. The authors concluded that the simulation package required to be combined with traditional teaching methods to enhance, rather than to replace. Henry and Holzemer[414] reached a similar conclusion following a study that compared clinical simulations with written methods of evaluation. Simulated activities have been claimed to be useful in a number of situations as shown in table 8.

Table 8: Indications for simulation, Neelamkavil, [415]

The real system does not exist
Experimentation in reality is dangerous or disruptive
The study involves past or future times, or when time requires to be compressed
Mathematical modelling of the system is impossible
Satisfactory validation of simulation models and results is possible.

It has been demonstrated that to be effective, simulated clinical problems must have certain characteristics [416] as shown in table 9.

Table 9: Characteristics of simulated clinical problems,[416]

The same information is used, as is available in the real setting
The situation must require a series of interdependent decisions
The simulation must provide realistic feedback regarding the consequences of the users action with each decision
Decisions must be final: student must experience the consequences
There should be several paths to the resolution of the problem

Miller [417] identified five levels of simulation fidelity: written, three-dimensional models, computer simulations, multimedia and simulated patients. The need for fidelity in a program varies with its purpose. Computer simulations have been useful for teaching theoretical problems whereas multimedia simulations and tutorials allows much higher fidelity, but at a much greater cost.

6.8.4 Learning theories and multimedia

It has been argued that a sound pedagogical approach to stimulate cognitive processes is required in the presentation of computer-based learning materials [337, 418]. However, the limitations of software and hardware have profoundly affected the pedagogical approach to learning that programmers have been able to use. Early CAL programs were characterised by linear progression and Skinnerean stimulus-response routines, simply because computer programming utilised 'if-then' logic that closely corresponded to characteristics of behaviourism [342]. Somekh [382] argued that it has been much more difficult to devise software that supports constructive learning, than to produce software designed on behaviourist principles. This could explain why behaviourism has become the dominant theoretical framework [418]. Early authoring tools did not allow for anything more imaginative[419]. For example, King[337] described a CAL development where cognitive strategies of concept mapping, chunking, framing and rehearsal could not be implemented using mainstream authoring software. Consequently, educationally

impoverished CAL programs that have been used and promoted in the past have perhaps soured both teachers' and students' views of the capabilities of computers in teaching. However, high powered authoring tools that allow branching and user control of routing through the application, such as Director™, Authorware™ and Toolbox™ have become more readily available, with the potential to produce software modelled on constructivist characteristics [418, 420]. These alternative approaches, although possible, require a new mindset by designers and developers that has often not been evidenced in the product of their activities[421].

The availability of powerful software has presented multimedia developers with choices and dilemmas in instructional design. For example, Rheaume[422] cited in King and Honeybone [423] stated that by browsing through multimedia, incidental learning is potentially greater than browsing a book or listening to a recording. Learning in this way can be enhanced by learners making links and connections for themselves, rather than following someone else's line of thought [424], which may lead to valuable insights [425].

Learning has also been demonstrated to be more efficient if a 'minimal instruction' approach is taken, allowing the user to explore and discover for themselves [403]. By contrast, offering nothing more than choices to the user has been described as a derogation of the teacher's responsibility to guide and direct students [426]. It is therefore an educational challenge to achieve an appropriate balance between giving guidance and allowing freedom within any multimedia program. An example of the type of balance that needs to be struck is in the provision of hypertext links. Hypertext presents a way to access on-line information that differs from reading text. Text is typically presented in a linear form, in which there is a single way to progress, starting at the beginning and reading to the end. Hypertext permits a reader to choose a path through the text unique to that individual and most relevant to his or her interests [427]. However, hypertext programs have the disadvantage of giving users the feeling of being lost[403, 425]. These conflicts are reminiscent of the tension between traditional and liberal schools of educational thought, as discussed earlier (3.4.2), right and wrong being a matter of perspective.

6.8.5 A pragmatic approach to instructional design

Learning theory permeates instructional design. Depending on the learners and the learning situation, different theories may apply. Some learning situations require highly prescriptive solutions, whereas others are more suited to learner control [428]. It has generally been agreed that learning is an active, constructive, cognitive and social process through which knowledge is constructed [408, 429]. In terms of instructional design, the skill and challenge is to determine under what circumstances behaviourism, cognitivism and constructivism ought to be used, how to achieve a pragmatic balance between these and how to engage the user in activities capable of supporting learning [342]. Instructional design is therefore a pragmatic application of theory, rather than a theory in itself.

Following comparisons of behaviourism, cognitivism and constructivism, Ertmer and Newby [430] concluded that instructional strategy and program content should depend on the level of the learners. For example, a behaviourist approach perfectly suited to novice learners might not be sufficiently stimulating for learners familiar with the subject content. This approach is consistent with that of Somekh[382] who argued that there is no universal right way of thinking about learning. However, many attempts have been made to link CAL and learning technologies to an educational paradigm. For example, Chandra et al [431] utilised the five varieties of learning and nine instructional events of Gagne[432] to develop Interactive Video programs. There is no consensus on this, however. Ohlsson[433] argued that embedding principles of cognitive theory into educational software was pointless because it was believed that cognitive principles could not apply consistently throughout, and would create unacceptable loss of flexibility in design.

Conole and Oliver [346] proposed a five-stage process neutral with respect to educational philosophies such as constructivism and behaviourism, which utilised Laurillard's[344] conversational framework as a model for educational interactions. This framework has been used successfully as a pragmatic structure for a simulation package for midwives [434], and for the pedagogical evaluation of virtual learning environments [398]. The creation of situations of cognitive conflict in the student (Laurillard's concept of intrinsic

feedback) was considered a motivational factor central to the learning experience. The differences in the how teachers and students experience and describe the world is vital for this. The teacher's goals and actions reflect the ways that a professional uses theoretical knowledge. The students, however, initially describe the world in simple terms in relation to the environment created by the teacher. Learning is the process, according to Laurillard[344] whereby the teacher and student engage in dialogue and feedback is given. In order that cognitive conflict is reduced, the teacher aims at helping the student acquire theoretical knowledge and become aware of the teacher's way of experiencing the world. This process is one of active engagement, resulting in transformation and construction of new knowledge.

6.8.6 Eclectic approaches

Somekh [382], drawing on the work of MacDonald[435] described a typology of computer interactions that is eclectic yet at the same time has been claimed to indicate levels of constructive understanding, as shown in appendix 7. The majority of software produced has been of types A, B and C, as types D and E have been limited by how well the computer can cope with open-ended free text entry [382]. As an alternative to considering the level of interactions, the level of functioning of the students themselves can be considered, and matched with appropriate theoretical approaches to learning. Jonassen[436] proposed a taxonomy, outlined in table 10.

Table 10: Jonassen's taxonomy of levels of learning related to instructional design[436].

LEVEL	DESCRIPTION
Introductory Learning	Learners have very little directly transferable prior knowledge about a skill or content area. They are at the initial stages of schema assembly and integration. At this stage classical instructional design is perhaps most suitable because it is predetermined, constrained, sequential and criterion-referenced.
Advanced Knowledge Acquisition	Follows introductory knowledge and precedes expert knowledge. At this point constructivist approaches could be introduced
Expertise	Is the final stage of knowledge acquisition. At this stage, the learner is able to make intelligent decisions within the learning environment. A constructivist approach would work well in this case.

It would thus seem important to identify different levels at which to pitch students' learning, as well as the context before any specific learning methodology could be recommended. In this regard Reigeluth's Elaboration Theory has been useful, as it organises instruction in increasing order of complexity, moving from didactic learning, to situations over which the learner has control [437]. This may work well in an eclectic approach to instructional design, since the learner can be introduced to the main concepts of a course and then move on to more student centred, self-directed learning that is meaningful to them in their particular context. Thus a behavioural approach could effectively facilitate mastery of the factual content of a profession (knowing what); cognitive strategies may then be deployed in problem-solving tactics where defined facts and rules are applied in unfamiliar situations (knowing how); then constructivist strategies, especially suited to dealing with ill-defined problems could be deployed, through (for example) reflection-in-action [430]. The suggestion that theoretical strategies can complement the learner's level of task knowledge allows the designer to make the best use of the available practical applications of different learning theories. With this approach, the instructional designer can draw on a large number of strategies to meet a variety of learning situations. This pragmatic approach could complement Goldstein and

Sorcher's training model [220], which uses a variety of approaches in the training of assessors of performance, as outlined in table 4.

6.8.7 Practical issues

The cycle of computer purchase, software development, obsolescence and redundancy has been a major dampening factor on the production of educational software [24]. In addition to hardware becoming obsolete, text-based, linear programs have been made redundant, as highly interactive multimedia products have become the norm in entertainment applications and games. It has been shown that regardless of content, a multimedia program will not achieve its educational aims unless it is of the highest production standard effectively creating and integrating media within the product [438, 439]. This requires much attention to be paid to principles of graphic design, and multimedia production. The fact that multimedia utilises sound, video, text, animation, graphics and photography, in an educational and interactive way means that skills in each of these areas are required [440, 441]. Few subject experts have these skills [440]. The formation of a team or partnership to include skills in project management, instructional design, graphic art, media, programming, and content, has been recommended [439, 442-444]. Schwirrian[172] suggested that expertise in content, technology and 'content-technology fusion' from the fields of psychology, education, engineering, sociology, art and information science is required. However, the reality of program development in higher education has been that colleagues and experts have been used to fill gaps in the knowledge and skills of the developer [445].

Another constraining factor on the development of multimedia is copyright. This can be a disincentive to the utilisation of existing media [446]. Permission from copyright holders is necessary when including media in software, and costs can be prohibitive. Bassett [447] cited the example of a two-minute clip from the popular 1960's cartoon 'Stingray' which cost £2,500.

6.8.8 Formative evaluation

The pragmatic nature of multimedia instructional design has made formative evaluation a necessity (*172, 402, 403, 448-453*), in order to ensure that the program has sufficient potential to achieve its desired outcomes. Formative evaluation contrasts with summative evaluation, the purpose of which is to measure the program outcomes against its stated aims (*453, 454*). There are good reasons for having a rigorous programme of formative evaluation in place during the development of a multimedia product; for example it can provide a business advantage, as costly errors and bugs can be identified early and their effects minimised (*399*). However, success in the marketplace, the distinguishing feature of 'good' commercial software (*399, 402*) is arguably not an appropriate test of the value of educational software as markets are small (*351*). For educational software, the only reliable gauge of quality is through a rigorous process of formative evaluation during production, followed by the summative evaluation of its educational effectiveness. In relation to this project, it was considered unethical to use a multimedia product on the experimental group before it had been rigorously tested and demonstrated its potential usefulness; formative evaluation served this purpose.

Various strategies have been used to evaluate software, such as Alpha and Beta testing. Alpha testing is conducted in-house towards the end of the production phase, to remove any major faults, whereas Beta testing provides potential users with access to the product, to fine-tune the application before mass production (*439*). However, Tessmer (*453*) argued that formative evaluation ought to be more than two stages: it ought to be an organised and precise process extending throughout the program production. Four stages of formative evaluation are described in table 11.

Table11: Tessmer's stages of Formative Evaluation(453)

EXPERT REVIEW
ONE-TO-ONE (SUBJECT WITH DEVELOPER)
SMALL GROUP EVALUATION
FIELD TRIALS.

These stages seem logical and imply a progression, with different approaches suited to different stages of program production. In addition to describing stages of formative evaluation, Tessmer identified five broad categories of information to be sought, as outlined in table 12.

Table 12: Categories of data gathered in Formative Evaluation: Tessmer[453]

LEARNING EFFECTIVENESS,
CONTENT QUALITY,
IMPLEMENTABILITY,
LEARNER INTEREST AND MOTIVATION
TECHNICAL QUALITY

Expert review:

Identifying actual "experts" who ought to be consulted can be problematic [453]. Selecting too many experts of the same type has been advised against because of the contradictions that can emerge [455]. Similarly Rushby[399] argued that control groups are not required, since the aim is to find out whether the aims of the program are achievable, not necessarily whether the process is better, faster, or cheaper than alternatives.

One-to-one

Subjects ought to be chosen to reflect differing abilities and experience to critique both the technology and the content [453]. The views of technically naïve have been considered essential to obtain, as computer literacy cannot be assumed in users [402]. Verbal Protocol Analysis is a common method of data collection and processing. During this process, each subject is asked to think aloud, as the program is being used. The researcher prompts the subject to explain what they are doing and why.. The transaction is recorded, transcribed, and analysed. This method has been used widely as a method of evaluating software [402, 403, 456, 457].

Small group evaluation

This involves a group of users in the evaluation of an unfinished version of a program [453], but can be distinguished from one-to-one by the use of the program in authentic physical and social contexts. It has been claimed to provide a more accurate measure of learner performance. A small number of subjects may suffice.

Field trials

These are situated evaluations, and can be used to confirm revisions, generate final suggestions, or investigate the effectiveness of the program [453]. In effect, the field trial in this research project developed to become the summative evaluation, which took the form of a Randomised Controlled Trial and follow-up telephone interviews. The rationale for this now follows.

6.8.9 Summative evaluation: The Randomised Controlled Trial (RCT)

The RCT research design has been considered the most rigorous way of determining whether causal relationships exist between an intervention and outcome [406]. At face value it would seem applicable to the evaluation of an educational multimedia program. However, problems with complexity, standardised procedures, sample size limitations, the issue of timing of effects, and ethical issues in relation to control groups have been shown to exist [404]. Widely used in clinical trials in medicine, it has been argued that doctors should only be involved in RCTs if it is believed that all treatments under study have potentially equal therapeutic effects [458]. It could equally be argued that the same ethical standards ought to apply to researchers in education. For a RCT to be considered ethically appropriate, firstly, there must be an indication that the intervention may be beneficial, and secondly, there must be recognised uncertainty over the effectiveness of the intervention [406]. The literature review has confirmed that these criteria apply in the context of developing a multimedia CAL program to help mentors assess student nurses.

The purpose of choosing a RCT for this study was to determine the effectiveness of the multimedia program as a teaching method, and to compare this with the traditional. However, as Jadad [459] explained, there are different types of RCT. Explanatory RCTs

are much stricter in their execution, setting rigid inclusion criteria to produce highly homogeneous groups. They also tend to use placebos or non-intervention control groups, use fixed regimes and collect hard, factual data as outcome measures, for example, marks in an objective test. Pragmatic RCTs, however, allow lax inclusion criteria, flexible regimes, active control groups and participants from varied backgrounds creating more heterogeneity within groups.

Given the context of this study, a pragmatic RCT design was chosen, as it was based in the real world rather than the sterility of a laboratory [460]. Therefore variation in size of study groups, time of day, control group teachers, and teaching sequence within the overall mentorship preparation programme, could be accommodated. Given the organisation of the mentorship programme, it would have been impractical to attempt to control all variables through inclusion criteria, however, through random assignment of individuals to comparison groups, known and unknown confounding influences such as gender and computer anxiety, ought to be distributed evenly between the groups ensuring that any difference in outcome is due to the intervention [461].

RCT follow up

Following up subjects in a RCT can enhance a study by providing qualitative data to help interpret quantitative results. It can also indicate delayed effects of the intervention, a recognised problem of the RCT [404]. For example, the RCT of Lo et al [386] compared the effectiveness of a CAL program with traditional teaching. Initially, no significant difference was found between experimental and controls, but after three months, motivation was higher in the experimental group. Follow-up could be beneficial if focused on the main areas of evaluation, or on issues that arise from the data analysis. In principle, therefore, follow up to the RCT was desirable.

As mentors can be distributed throughout a large geographical area, an efficient, but effective method of follow-up was required. Telephone interviewing can be a convenient method of collecting data, especially if the interview is short, specific and impersonal, or

if the researcher has had prior contact with the subjects[2]. These criteria apply in this research.

Structured interviews gathering quantitative data using a schedule can gather information in a standardised way; this has been the most common use of telephone interviews. A criticism, however, is that this process may not reflect the diversity of responses [462]. Qualitative research using the telephone, although much less common, is capable of being performed efficiently and successfully, and can yield data rich in detail [463]. Another disadvantage is that interviewing can compromise the validity and reliability of the generated data. To explain, the more the interviewer attempts to increase reliability by becoming rational, and detached, the less likely the interview is perceived as a friendly discussion, and the more calculated (and less valid) is the response from the subject [464]. A careful balance therefore needs to be struck. Questions need to be phrased in a standardised way, but open-ended to allow sufficient breadth of responses. One advantage of exploratory, focussed interviews is that exact representativeness or large sample size is not crucial [272].

6.9 Conclusion

It has been recognised that both cognitive and behaviourist approaches have their place in computer assisted learning and multimedia type programs, and can be used in complementary ways [342, 425, 465]. However, it could be argued that using the Internet to learn and web-based courses are more appropriate as computer mediated communication enables the teacher to respond to student input, no matter what the content of that input might be. Dialogue in the context of stand alone cal programs is not verbal communication in the same sense, but occurs through cognitive interaction with content, which in turn is a function of the program's instructional design. Using an appropriate instructional design, governed in part by formative evaluation, ought to be capable of producing a program that achieves real learning for mentors, measurable in a pragmatic randomised controlled trial. The aims of this research, explicated from the literature review, are outlined next.

Chapter 7: Aims, Objectives and Hypotheses

7.0 Summary

This chapter states two aims and associated objectives and hypotheses of this study. The first is in relation to the design and production of the multimedia program; the second is in relation to the methods and processes of the evaluation of the educational effectiveness of the program.

7.1 Aim 1

To develop and formatively evaluate the multimedia program 'A Mentor's Guide to Assessing Student Nurses', that facilitates the acquisition of skills in the observational assessment of student nurses' performance.

Objectives were to

7.1.1 Model and evaluate the process of clinical assessment to facilitate the on-screen exploration of underpinning theory.

7.1.2 Demonstrate key assessment concepts, select and evaluate the instructional design strategies that facilitate their understanding on-screen.

7.1.3 Simulate the assessment process on-screen selecting and evaluating the instructional design strategies to produce:

task diagnostic feedback

verbal encouragement

role modelling

vicarious learning

a psychologically safe environment for learning

7.2 Aim 2

To evaluate the educational effectiveness of the program through comparing the multimedia method with the traditional teaching method in a randomised controlled trial.

Objectives were to

7.2.1 Construct and evaluate a questionnaire measuring anxiety, attitude, self-efficacy and knowledge to indicate mentors' assessment potential by determining:

7.2.1.1 face validity with lecturers and mentors

7.2.1.2 concurrent validity through interviewing lecturers experienced in supporting mentors and supervising the clinical experience of students

7.2.1.3 test re-test reliability: correlation co-efficients >0.70

7.2.1.4 internal consistency: alpha >0.70

7.2.2 Analyse demographic data to detect any systematic differences between study groups and the population of mentors at large, by testing the following null hypotheses:

7.2.2.1 that 'there is no significant difference between control, experimental, and mentor database groups with regard to:

Place of employment
Gender
Clinical grade'

7.2.2.2 that 'there is no significant difference between study groups with regard to

Number of years qualified
Prior experience of assessment
Educational background
Professional qualifications held
Frequency of exposure to students in the workplace
Baseline scores for
anxiety,
attitude
self-efficacy.'

7.2.3 Analyse data from the 'assessment potential' questionnaire in the Randomised Controlled Trial to test the following null hypotheses:

7.2.3.1 'there is no significant difference between normative anxiety data, and the baseline measurement of anxiety in the RCT study groups.

7.2.3.2 'there is no statistically significant difference between the multimedia and traditional teaching methods for improvement (from before to after the session) in:

anxiety

attitude

self-efficacy'.

7.2.4 Explain extreme or unexpected responses to multimedia and traditional teaching methods through focussed telephone interviews.

Chapter 8: Program Development

8.0 Summary

This chapter gives an overview of the development project, re-capping on the program's rationale, conceptual approach and aims, before describing issues in its development, such as costs, copyright, and ethics. The instructional design of each part of one version of the program, appendix 1b, is described, with rationale derived from the literature review. The program was constructed using behaviourist, cognitivist and constructivist methods in three main conceptual areas, 'The Assessment Process', 'Concept Demonstration', and 'Assessment Simulation'. Case studies and video of real students and patients on location have increased the fidelity of the learning experience to enhance transfer of learning. A strategy of formative evaluation of an early prototype of the program (appendix 1a), including expert review, small group evaluation, and verbal protocol analysis, guided the instructional design. Findings were mainly positive, with criticism in the areas of technical quality, navigation, and user guidance. Conclusions and recommendations allowed refinement of the program (appendix 1b), which was then of sufficient quality to be used in a randomised controlled trial without ethical constraint. The production of a PC compatible version (appendix 1c) marked the completion of the program's development.

8.1 Development project overview

The literature review indicated what the aims of the program should be and how these could be achieved using a multimedia program. Factors that have been shown capable of affecting a mentor's ability to assess students were identified in the literature (5.5-5.6, pgs 66,72). Mentoring has been difficult and uncertain, resulting in anxiety, negative attitudes and reduced self-efficacy in relation to the assessment of students' practise. The computer program was therefore designed to improve mentors' knowledge and understanding, improve attitudes, reduce anxiety, and improve self-efficacy by providing information, motivation, opportunities to practice, and feedback on performance. Authorware™ was purchased and a partnership formed with a multimedia designer. The

process of program development took approximately 18 months, with cycles of content research, media production, formative evaluation, and program assembly. Integral to this was the necessity to develop an appropriate instructional design, with strategies that complemented the message, the medium, and the sophistication of the learner.

The overall structure of the program comprised three parts, to reflect each of the objectives (7.1.1 – 7.1.3) of Aim 1. Each part was quite different in instructional approach involving behaviourist, cognitive, and constructivist techniques. Additional, peripheral functions were developed, such as the guided tour and the development of a 'Mentor Support' web-site. The most time consuming and labour intensive part of the process of developing the program was media acquisition. Filming involved many individuals (appendix 8) for whom scripts had to be written and storyboards prepared.

Production was time consuming, with little opportunity to recoup costs. An abortive attempt to collaborate with publishers Churchill Livingstone (appendix 9) made this apparent. Initial interest in the project waned because of uncertain market conditions. The effect of this was constant pressure to be economical with both time and money, which had consequences for quality, especially video. Through supportive management, funds were acquired and the necessary hardware and software purchased. Audio-visual technicians, a video-editing suite and the use of a sound recording studio were provided. As video files can be very large, it was planned to create the program for CD-ROM, allowing the program to be 650 MB, maximum. This constraint affected quality, as the quality of data and audio files had to be reduced in order to be accommodated, even with optimal data compression. The sentiment of Laurillard [426]not to allow limitations on technology to influence instructional design denies these facts of life, and instructional design needs to be pragmatic.

Copyright

Run-time only programs in Authorware™ were free of copyright restrictions when distributed for academic purposes, as too was the library of clip-art drawings, backgrounds, and icons purchased for the project.

Project management.

A technician with qualifications in graphic design and skills in multimedia authoring and programming was recruited to the project enabling program authoring to be at a quicker pace and to a higher standard than would otherwise have been possible. Coding, digitising video, and manipulating photographic images can be complex activities with steep learning curves.

Context of learning

The program was designed to be distributed to potential mentors, who would then have access to high quality learning resources in the workplace. However, it was also designed to accommodate a range of teaching situations, for example, to prepare student nurses for an assessable placement. In addition, the program could be used by small groups of individuals rather than one individual at a time, up to a maximum of three. The impact of feedback from peers has been shown to be much more powerful than that automatically generated by computer [409]. Finally, the program could be used with a larger group using a data projector in a classroom or lecture theatre setting. These potential uses indicated that an easily navigable program structure was required, with discrete units that could be accessed rapidly.

8.2 Ethical considerations and ethics approval

Ethical issues in relation to this research were discussed initially with the research supervisors and then with the Principal of the College of Nursing and Midwifery, in the absence of an ethics committee. The outcome was the identification of issues worthy of consideration, such as full disclosure of the purpose of the study and informed consent from all involved. It was thought that where lecturers and professional nurses were being asked to feature in video or review learning material, then tacit consent could be assumed from their participation. However, where students or patients were included in filming, then both verbal and written information ought to be given and written consent obtained in recognition of their vulnerability. Moreover, because of the invasive nature of filming in ward environments, institutional ethics committee approval was required. Application

was made to the ethics committee of one local NHS Trust to allow the participation of staff and patients in this project, including filming in clinical areas. Consent forms and information sheets were included, as shown in appendices 10 –12. Following a meeting with a sub-group of the committee, approval was granted (appendix 13).

As part of the process of asking for participation, the purpose of the research was explained. Whereas research data was stored in accordance with standard protocols, made anonymous where possible and destroyed after processing, this was not possible with video, audio, or photographs and certain safeguards were necessary. Because video recordings cannot be anonymous to anyone who knows the participant, it was not undertaken lightly. Students were filmed for approximately four hours each, in medical or surgical wards, to gather sufficient footage from a range of activities representing each of Benner's Domains (table 2). Consent forms were signed (appendix 11), and editorial control agreed. Patient participation followed informed consent: information sheets (appendix 12) were distributed the previous day, and relatives had an opportunity to discuss the project. This process was consistent with ethics committee approval (appendix 13).

When media were edited, digitised and included in the program, participants were contacted and invited to see how their images were being used. Signed patient and student consent forms have been retained for as long as the media continue to be used and participants have the right to withdraw their images and audio recording from the program at any time. Video in excess of requirements was destroyed.

8.3 'A Mentor's Guide to Assessing Student Nurses': design, interface, and navigation.

This part illustrates and focuses on the details of the program, appendix 1b. This was because this version was the product of formative evaluation and used in the subsequent Randomised Controlled Trial. A holistic approach to program design was taken, which concentrates on the visual appearance of the user interface and its functionality [403]. Careful attention was paid to principles of human-computer interaction and instructional design, as well as to the appropriateness and accuracy of content.

Navigation

The process of navigating through a computer program can be daunting. A useful distinction can be made between *conceptual* navigation through the ideas the content represents and *physical* navigation between the media components [403]. The content of the program indicated a logical division into three parts (table 13)

Table13: Parts of 'A Mentor's Guide to Assessing Student Nurses'

PROGRAM PART	CONTENT
The Assessment Process	An outline of the steps involved using a method of navigation that promotes user control, to achieve objective7.1.1.
Concept Demonstration	An interactive demonstration of key observational assessment concepts: selecting a sample and judging performance against set criteria. Feedback on understanding is provided. This will achieve objective 7.1.2.
Assessment Simulation	A representation of the main elements of clinical skills assessment, simulating the process of gathering and organising data, and arriving at a judgement of performance. Feedback on the user's ability and progress in making a satisfactory judgement of the student is provided. This will achieve objective 7.1.3.

An open, freely navigable structure was desirable, in order that users could quickly identify and move to the area most likely to meet their learning needs. Therefore, a structure that encouraged 'browsing' was developed. Only a small number of linear sequences were necessary, for example, text pages. There are three occasions where a specific type of input from the user is mandatory, and for which the computer waits. This is during the 'data entry' in the 'self-assessment' and 'simulation' parts of the program:

once begun, the user needs to complete the data entry and receive feedback before moving on.

Guidance instructions

A minimal instruction approach was taken, to encourage the user to think about their learning needs and which parts of the program might be most helpful. However, some orientation was necessary to avoid a 'discovery learning' approach that can cause frustration and inefficiency. This program aimed at striking a balance by providing a freely navigable interface with low-key, non-mandatory direction through:

- The guided tour
- Text introductions to sections
- The use of an advanced planner in the 'Assessment Process'
- Directional arrows to sequence data collection in the simulation
- Feedback statements in the simulation

In order that the user quickly learns to navigate the program, consistency of screen design has been adopted as a strategy, with identical control widgets and menu icons on each page. From any point within the program, the user can click on any of the familiar menu icons to return to the beginning of a major program part. Because metaphors and icons are heavily culture dependent [403], care has also been taken to reflect nursing as an independent profession within a modern, multi-cultural society.

Types of media

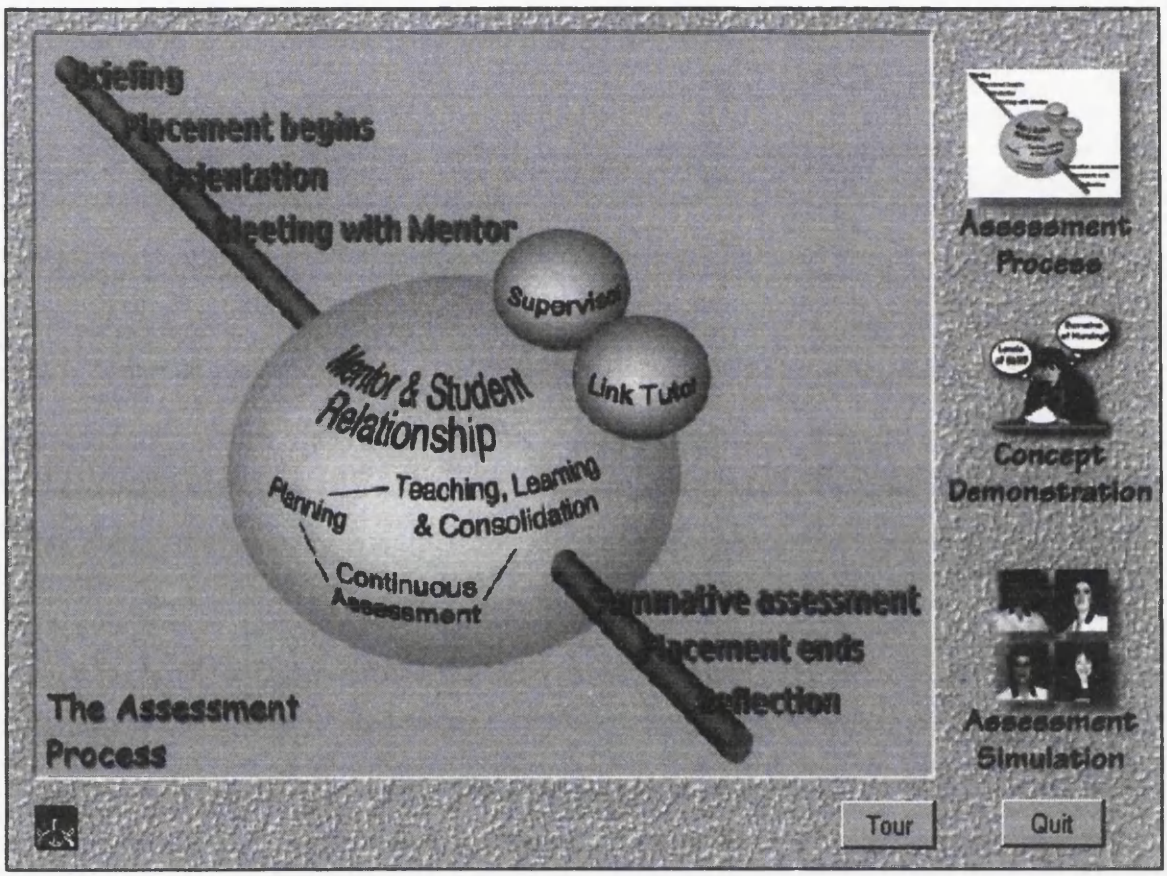
Consistency of approach can be argued to be beneficial[403], so too can a variety of media to maintain interest and motivation. Different media, including video, photographs, animation, text and audio have been selected for their strengths, and used in balance throughout the program.

8.3.1 'The Assessment Process'

Overview

This part of the program has been designed to allow the user to examine inter-related stages of the assessment process, with easy access to explanations of the underpinning theory. A three dimensional graphic with hypertext links allows the user to see the context of each component in the process, which can be explored in greater depth, as shown in figure 2.

Figure 2: Screen representation of 'The Assessment Process'



Instructional design

The 3-D graphic is an 'advance planner' of the content of this part of the program. Steps in chronological order have been identified, from the beginning of the placement to the end. The intention has been to encourage hypertext browsing where the user investigates areas of interest, rather than attempting to read everything at once. Text was presented on

sequential rather than scrolling pages, as this has been claimed to be superior in situations where the reader has to organise, understand, and integrate content [425].

The graphic was constructed and illustrated with reference to models of the assessment process discussed in the literature review (4.8, p53). Journal articles and books in support of the theoretical concepts under discussion that may be of interest to mentors have been included in an accompanying 'Mentor Support' web-site, under 'Resources'.

8.3.2 'Concept Demonstration'

Overview

This part of the program demonstrates two areas that mentors ought to understand when assessing students: 'sampling performance' and 'judging performance'. 'Sampling' aspects of performance allows a user to practice and become familiar with Benner's Domains of Nursing Practice [90]. These were outlined in table 2, and have featured strongly in the literature review, parts 3.5.1 (p30), 3.8.8 (p40), and 4.6.3 (p49), as an acceptable framework for dividing performance into discrete aspects for independent judgement. 'Judging' performance examines the differences between levels of performance by taking the example of one domain, judged using the complementary conceptual frameworks of Bondy (appendix 4) and Steinaker and Bell (table 3).

Instructional design

The content indicated that a sub-division was required. The main concepts, entitled 'Sampling Performance' and 'Judging Performance', have been closely linked. However, in order that details can be explored they have been treated separately in this part of the program. The explicit deconstruction of an activity has been identified as an effective instructional design principle[433].

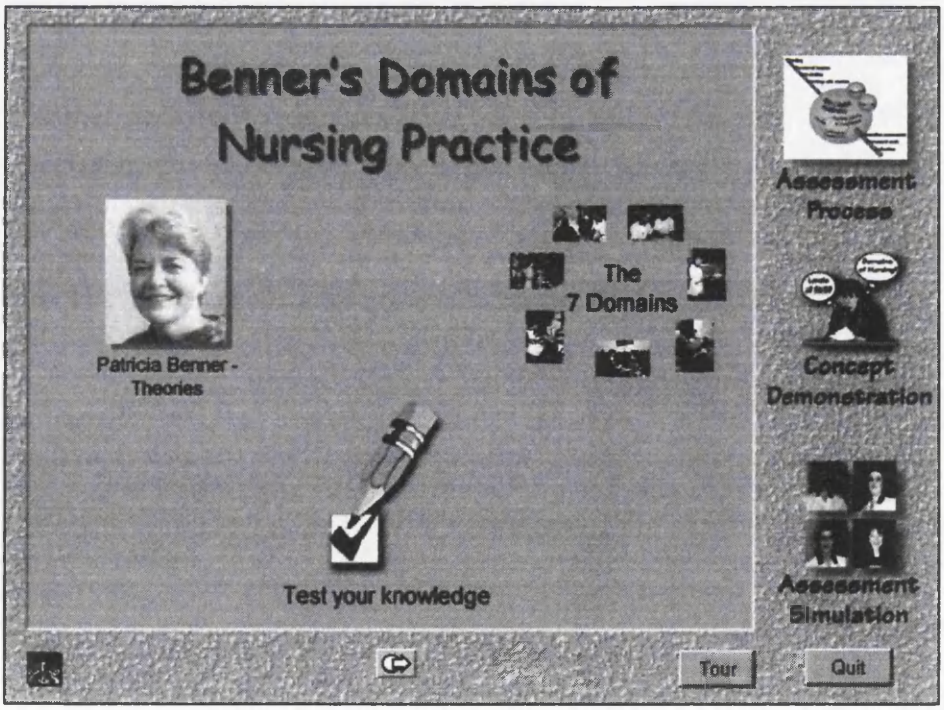
Video has been chosen as the main medium to present complex ward and community based situations as it has been found useful in training observational assessors of performance (4.9.3, p56), in teaching the social sciences, and in situations where real

tasks have to be closely observed [425]. The intention was to give users the opportunity to put their understanding of concepts into practice, by identifying Benner's Domains and judging different levels of skill as portrayed in short video clips. Feedback on success or failure has been aimed at providing both reinforcement, and the cognitive conflict necessary to re-consider the issue. Feedback statements were vetted by an expert panel during the process of formative evaluation of the program. The behaviourist technique of drill and practice was adopted, as this has been shown useful for introductory learners (table 10) and effective for training purposes [466].

'Sampling Performance'

Sequential text pages briefly explain the origins of Benner's Domains and their relevance to the observational assessment of performance. A sub-menu, as shown in figure 3, allows the user to discover more about each domain, with opportunities to demonstrate knowledge and understanding through the correct identification of the domain in a number of short video clips. Feedback on both correct and incorrect choices has been provided. Video controls and sliders allow the user to stop, repeat or end video clips as desired.

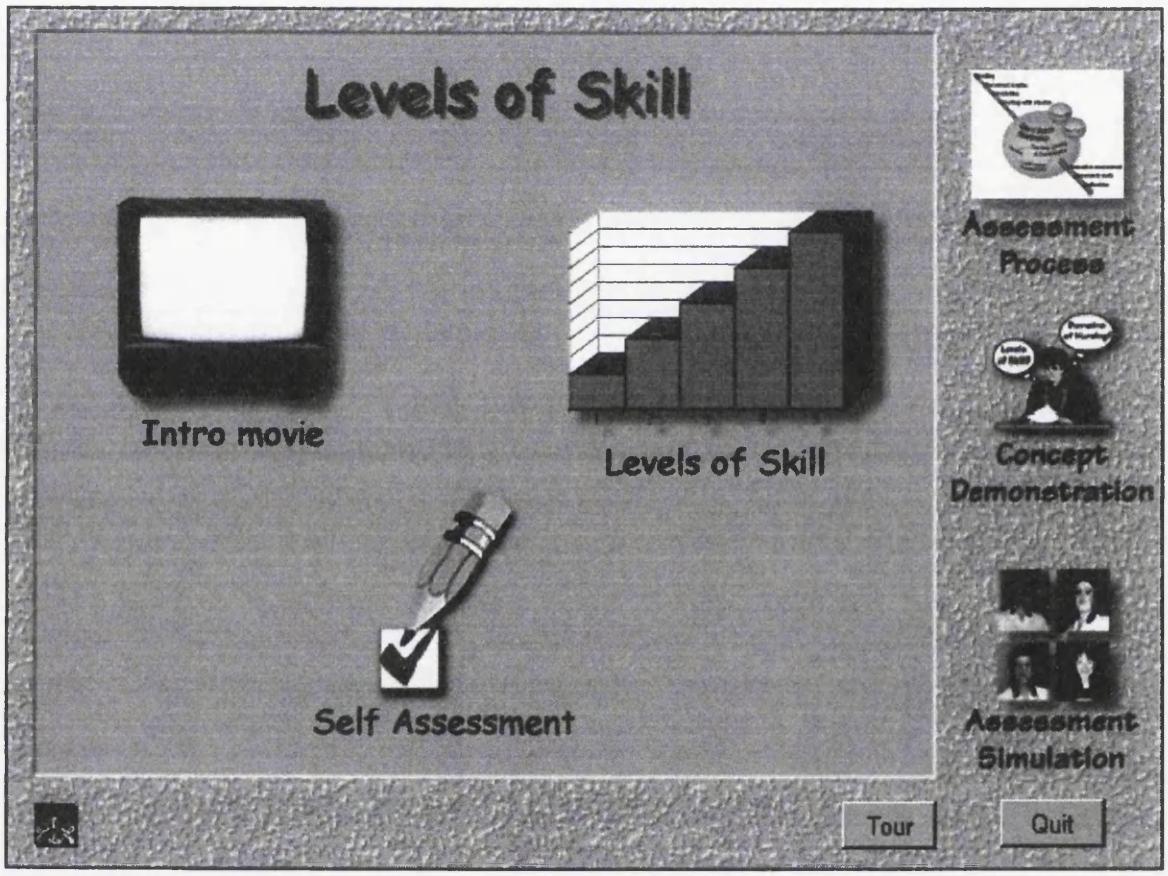
Figure 3: sub-menu within 'Sampling Performance'



'Judging Performance'

This focuses on the concept of judging levels of performance. Sequential text screens with embedded hypertext links provide an introduction. A sub-menu follows, as shown in figure 4, inviting the user to click on the 'Levels of Skill' icon for further exploration, or an introductory movie to the scenario, or to test knowledge using video clips, as in 'Sampling Performance'.

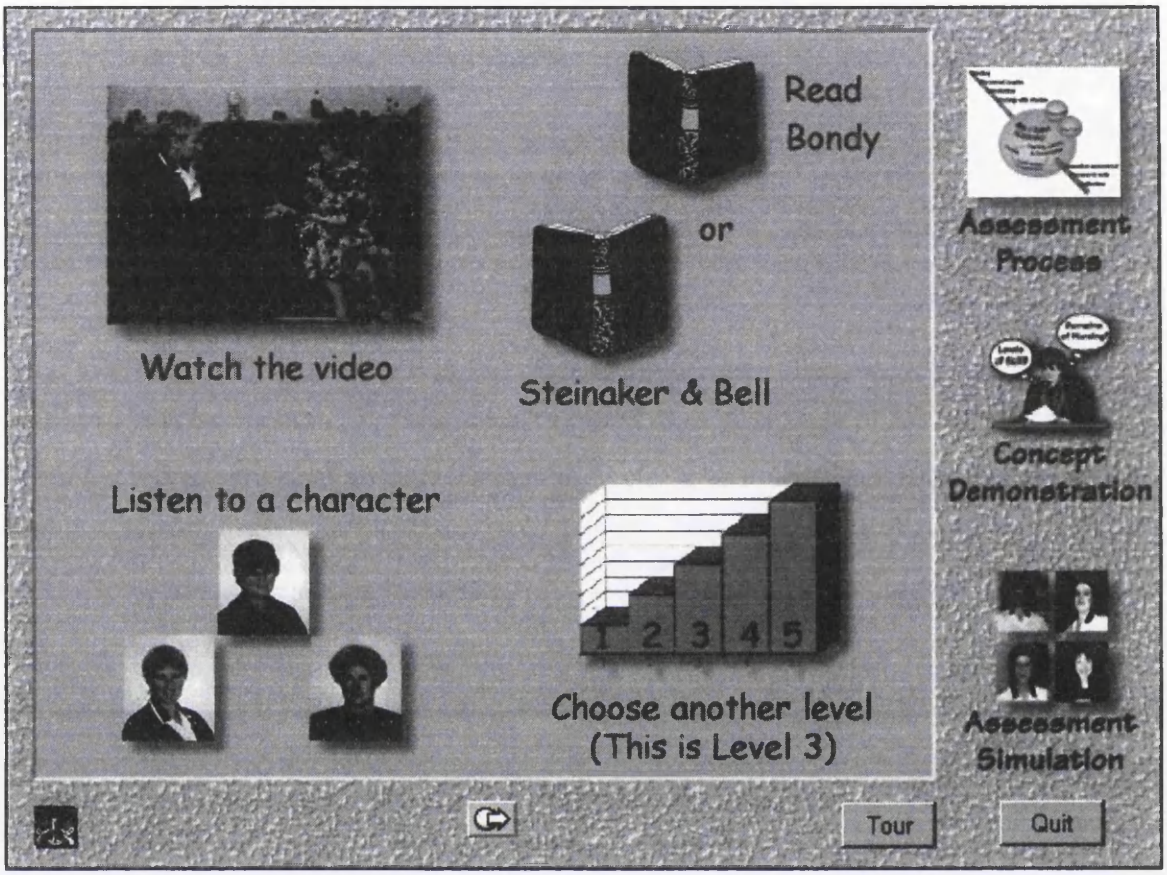
Figure 4: Sub-menu within 'Judging Performance'



Pointing and clicking on 'Levels of Skill' leads to one further sub-menu. The user is invited to select one level of skill from a possible five. For each level, the user can explore both Bondy and Steinaker and Bell's terminology and characteristics of student performance, with an indication of what the role of the mentor might be. For each level, the user is presented with a screen as in figure 5, allowing the selection of video, text and

audio, centred on the scenario of a student performing within Benner's 'Teaching and Coaching' Domain.

Figure 5: Sub-menu: 'Judging Performance'



The context, actors, and situation have been controlled in each of the five levels, the variables being the performance of the student and reciprocal role of the mentor. Also available to the user are reflective comments by each of the main characters within the scenario.

The video for this part of the program was acquired in a community location, using amateur actors, under the direction of the researcher. Scripts were written in collaboration with diabetic liaison and district nurses to ensure accuracy and currency of information.

8.3.3 'Assessment Simulation'

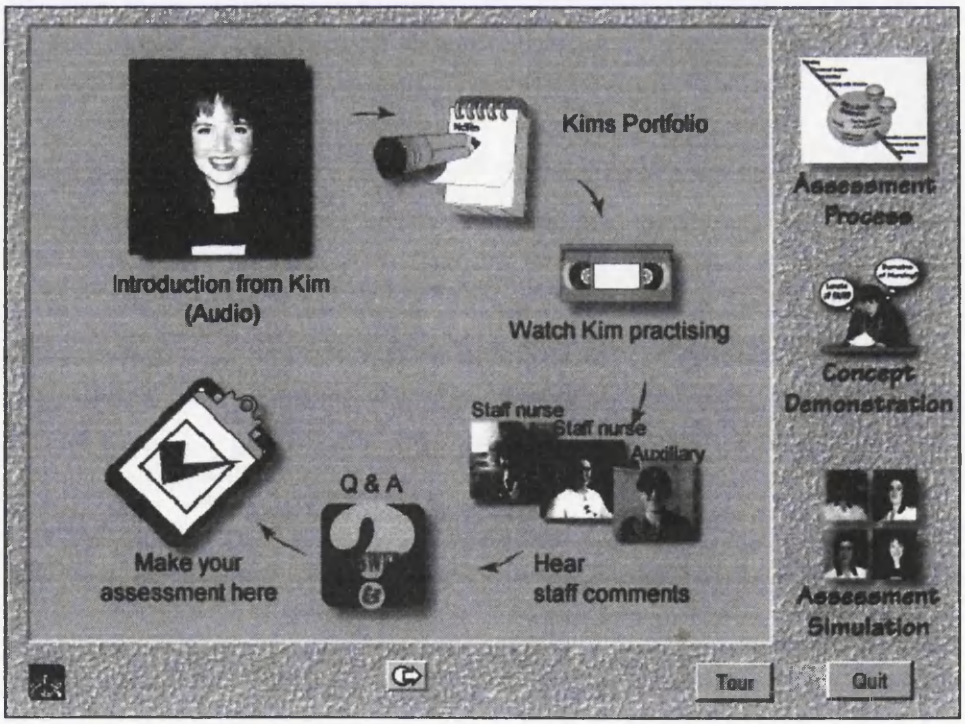
Overview

Following a text page introducing this part of the program, users may select from four student 'case studies'. This part of the program attempts to create realistic situations by integrating 'Sampling Performance' and 'Judging Performance'. The user is expected to choose relevant information, categorise it into domains, weigh conflicting evidence, and judge the student's performance, which is highly intense cognitive activity. Through providing feedback, the user is encouraged to improve upon past performance. The educational approach is consistent with the philosophy of problem based learning, involving the essential processes of exploring, analysing and deciding [467].

Instructional design: accessing information

This section has been designed to be non-linear, but the most logical route, and that most likely to produce cognitive engagement has been indicated to the user by arrows in the menu screen (figure 6). Subtlety was thought preferable to imposing one route upon the user and is in keeping with constructivist rather than behaviourist approaches to learning.

Figure 6: Simulation sub-menu



The screen introducing the simulation gives advice to the user to take notes. Information is available from the following sources:

- Student introduction: an audio recording made accessible through pointing and clicking on the student's photograph.
- Video evidence of the student's performance: made accessible through point and click on the video icon. Controls at the foot of the video screen allow fast forward, pause, and rewind, thus allowing the user to control the duration of the video, each of which can be as long as 8 minutes.
- Verbal evidence from staff: audio recordings made accessible through pointing and clicking on any of three staff photographs.
- Hypothetical questions and students' answers: made accessible through point and click on the Q and A (questions and answers) Icon. This has been organised according to the Domains of Nursing Practice to which the question relates.

The validity of an assessment is improved when the 'domain of content' of the student's performance has been adequately sampled (4.5.3, p45). Therefore, no single source of information was sufficient to judge the student completely. For example, organisational and work role competencies were difficult to encapsulate within a few minutes of video. An accurate assessment of performance in this domain therefore needs to be inferred from a number of sources, including the student's responses to hypothetical questions, and from the views of members of staff. This mirrors reality, where inferences are made and contributions from other members of the ward team are expected. Some conflicting information, for example, the views of different members of staff, was included in the program to prompt the user to evaluate data based on supporting evidence. Again, this mirrored reality and was intended to motivate the user to learn.

Instructional design: data entry, data processing, and feedback

The method of entering data to the assessment grid has been included in the 'Guided Tour', and described briefly in the introduction to the 'Simulation'. Hypertext links provide the user with access to short definitions, by pointing and clicking on the domain name, and the number of each level. Definition windows disappear by pointing and clicking.

Data entry and feedback

Users are expected to enter their judgement of the student's performance in each of the 'Domains of Nursing' in turn, by pointing and clicking in the appropriate box. This is an irreversible linear sequence, however, should a mistake be made, the user can return to this part of the simulation and re-enter data. Details of the logic by which this information is processed, analysed, and feedback produced is given in appendix 14. Essentially the computer compares the user's judgement with that of an expert panel who had been presented with the same information during formative evaluation (see 8.4.4, b, p132). An example of a feedback statement is given in figure 7. On the next screen, an indication of progression is shown on an animated scale, the pointers indicating current score and previous score, as demonstrated in figure 8.

Figure 7: Feedback statement example

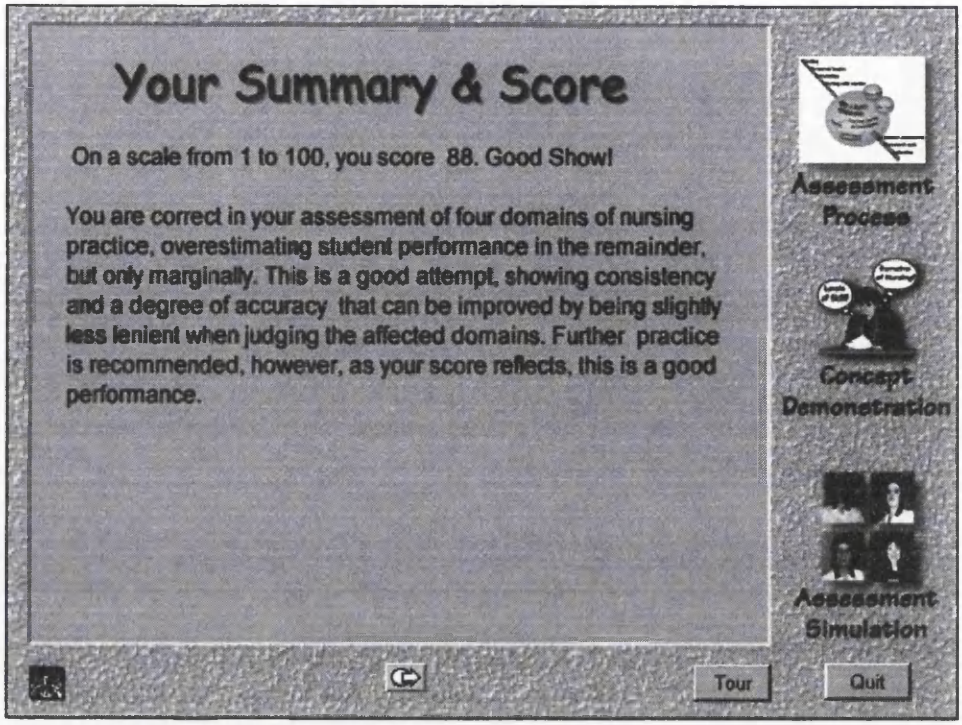
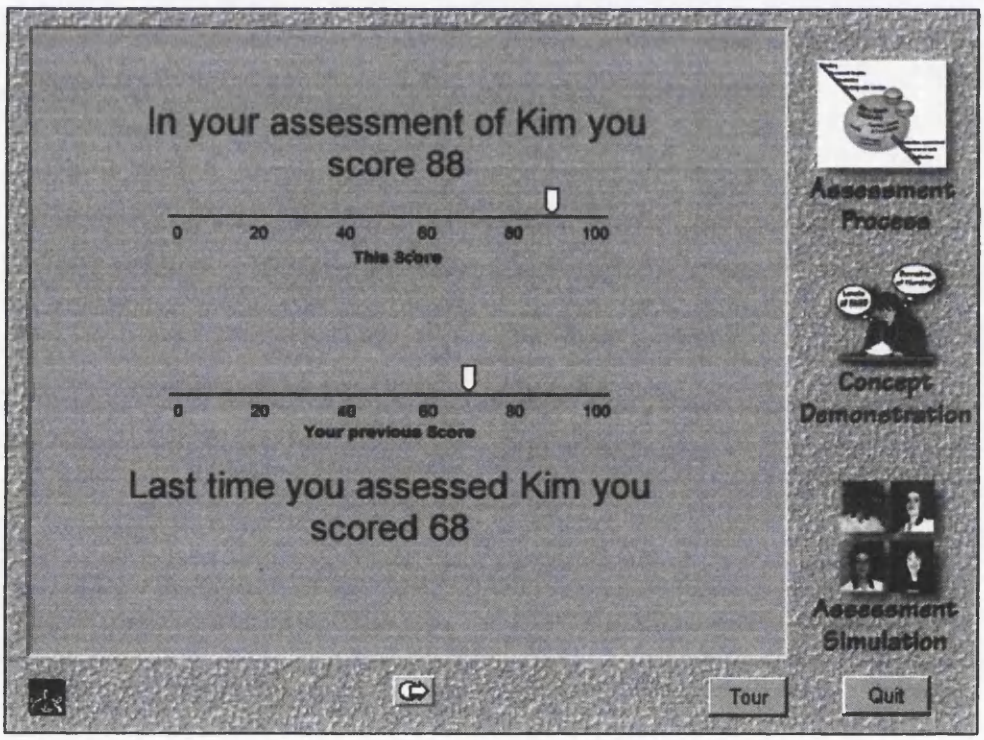


Figure 8: Comparison of current with previous score



Justification for this feedback design

Through these mechanisms, analysis of input, interpretation, advice and direction may be given to the user. The user is also provided with a sense of progression, an important factor in improving skills and knowledge [409]. This feedback design evolved because it focused on the judgement of the user, not the content of the case study. Any number of case studies may therefore be added to the program with no extra coding. All that need be entered are the values for the 'expert panel' assessment, i.e. the variables e1-e6 (see appendix 14). An alternative approach would have required the analysis of content and the creation of feedback statements specific to all possible user entries, for each additional case study included, making this impractical.

One disadvantage could be that feedback on the user's judgement in a particular domain, that is, whether or not a particular domain was scored correctly, is not given. The intention was to stimulate curiosity and cognitive conflict. The natural desire to achieve an improved score was considered a motivating factor and would encourage the re-assessment of the same student. This would be advantageous to learning. The alternative, to give the correct score, would have been a disincentive to continue to explore the content of each case study in greater depth. This issue has been discussed further in the formative evaluation of the program.

8.3.4 'Log-on'

This requires the user to enter a name and password before accessing the program. For each new user, the program produces a file on the hard disc, which is opened each time the user subsequently logs onto the program. This file stores the user's ALHR scores (appendix 14) generated in the simulation part of the program, accessed by the computer under password protection, maintaining confidentiality.

8.3.5 'Guided Tour'

This has been included in order to provide an overview of the program for first time users. As the program was non-linear, the user required some knowledge of the structure of the program, in order to make an informed choice as to where to begin. Users may be time limited, and require to focus on only one particular part of the program. This facility enables users to make decisions about which parts of the program to access, and in which order.

Instructional design

This was a linear, fast-paced combination of audio, text, animation, graphics, screenshots and images that lasts approximately four minutes. No input is required from the user once the guided tour is launched, however its use has been made optional, as it has been designed for first time users and for those having major difficulty with navigation. The overall aims of the program have been included, followed by a description of the program and its component parts. This was scripted and the soundtrack created in a recording studio.

8.3.6 Summary

The program "Assessing Student Nurses" was developed using a pragmatic mix of behaviourist, constructivist and cognitivist methods. Fidelity was achieved through using real students and real patients, in real situations, to optimise the transfer of learning. The program provides the theoretical underpinning of the process of assessment, creates opportunities for practise, and provides feedback on performance using an intuitive, icon-based interface. This was achieved using a liberal, student centred approach consistent with other aspects of nursing curricula post-Project 2000. The next section describes how this instructional design evolved through the process of formative evaluation.

8.4 Formative evaluation of "A Mentor's Guide to Assessing Student Nurses"

8.4.1 Introduction

The purpose of formative evaluation was to ensure the development of a quality educational product, the effectiveness of which could then be gauged in a Randomised Controlled Trial. Tessmer's stages of formative evaluation, as described in table 11, were used as a conceptual framework. The first three stages were fully concerned with formative evaluation, initially involving alpha and then beta versions of the program. The last stage, 'field trials' evolved to become the Randomised Controlled Trial itself.

In addition to describing stages of formative evaluation, Tessmer's five broad categories of information, described in table 12, were useful in formulating appropriate questions to ask, and in selecting individuals who were best placed to provide the type of information required. What follows is an account of the process of undertaking the systematic formative evaluation of an early prototype version of 'Assessing Student Nurses' (appendix 1a), that resulted in recommendations (such as changing the title to 'A Mentors Guide to Assessing Student Nurses') that were incorporated into appendix 1b, the details of which have been described (part 8.3, p116).

8.4. 2 Stage 1: 'Expert Review'

In practice, one or two individuals can be sufficient for an expert review, as explained in part 6.8.8 (p105), however, more than this was necessary, in order to address content and pedagogy. The process began with an interview with a professor of education with expertise and a record of accomplishment in mentorship and educational research. The discussion focused on the concept of competence and the assessment of practice in the professions, brainstorming ideas for the program, which at that time was in the initial stages of development. The outcome was useful in identifying the main concepts to consider.

Individuals with knowledge and expertise in multimedia were also utilised, including a computing consultant, an educational psychologist specialising in the evaluation of CAL, a multimedia expert with a special interest in software production, a publisher who

developed and marketed educational multimedia software and two senior lecturers with expertise in CAL and Health Care Informatics. An early version of the program was used with these individuals and notes taken; or left with them, with a request to comment critically on each of the five Tessmer categories (table 12).

In addition, the Scottish Interactive Technology Centre, Edinburgh, and the Scottish Council for Educational Technology, Glasgow, were visited with the purpose of reviewing methods of programming from pedagogical and design perspectives. Less formal, local expertise was utilised to review content and to proof read text for inclusion in the program. One Lecturer and one Senior Lecturer were willing to undertake this task. The fidelity and accuracy of filmed situations and verification of the different levels of skill demonstrated by video were ascertained by community nurses reviewing scripts and analysing video during editing.

8.4.3 Stage 2: 'One-to-One'

Four nurse lecturers experienced in teaching and supporting mentors were asked to review an early version of the program individually, screen by screen. These subjects were chosen to reflect differing abilities and experience of CAL and multimedia. Verbal Protocol Analysis was used. Each expert was interviewed and recorded while using the program, using a semi-structured interview format and a schedule of questions as shown in appendix 15. Answers to selected questions were sought during screen by screen review of the program. This ensured that all content was checked, and that the main routes through the program were followed. Complex areas thought likely to present difficulties, such as the simulation, were given particular attention. In addition to notes taken at the time, the answers to questions and spontaneous comment were recorded for later transcription and analysis.

Subject expert 1

This lecturer was experienced in word processing, spreadsheets and databases; had studied Mathematics at Degree Level and had a background in computer programming. It was considered unlikely that this individual would be daunted by the program. The ability

to give a true evaluation free from relative influences is important, as this can undermine the value of experts' views[374]. This individual was the co-ordinator for mentor preparation in Glasgow, and was considered capable of focusing on the content of the program and judging whether the material was accurate and sufficient to meet the anticipated needs of trainee mentors.

Subject expert 2

This lecturer had an interest in CAL and was experienced in its use. He was thought capable of criticising the program from practical and pedagogical perspectives.

Subject experts 3 and 4

These individuals were similar in that they were very experienced nurse educators but had rarely used computers. Booth [402] suggested that such 'naïve users' might reveal difficulties that computer literate individuals subconsciously compensate for [399]. These experts were thought capable of identifying fundamental errors and misunderstandings within the program.

8.4.4 Stage 3: 'Small Group and Peer Evaluation'

The most convenient group were mentors who were undergoing or had recently undergone a programme of mentor preparation. A small number from one location was used to minimise contamination of the population from which the sample for summative evaluation would be drawn (see 9.2.4, p159).

(a) Staff nurses undergoing mentorship preparation

A group of three prospective mentors were discretely observed and recorded while using the program. Generally the mentors were very impressed and found the program motivating and easy to use. Good quality debate occurred within the group, especially when the performances of students within the case studies were analysed.

(b) Mentors from a local hospital 'Mentor Support Group'

Mentors from this group were asked to form an expert panel and review the material presented in the 'Simulation', and self-test areas of the 'Concept Demonstration' part of the program. The 'correct' response against which the computer gauged the user's input was reached through consensus, encoded and incorporated into the software.

(c) Presentations to professional and academic colleagues

During the course of production, obtaining the views of professional and academic colleagues was necessary in order that the program gained acceptance. A strategy to identify opportunities for presenting the project at gatherings both locally and nationally was adopted, involving seminars and conference presentations.

Seminar presentations

There were opportunities for five seminar presentations. Two at Glasgow College of Nursing and Midwifery, one at Glasgow Caledonian University, one at Glasgow University and one at the Victoria Infirmary NHS Trust, Glasgow. Attendance ranged from five to thirty-two. These groups were useful for encouraging critical analysis of the project's aims and direction, appraising the content of the multimedia program, and generating ideas for its evaluation.

National conference presentation

In addition to local seminars, abstracts were submitted for larger gatherings of nurses and nurse educators. The first of these was a National Board of Nursing, Midwifery and Health Visiting for Scotland conference on Mentorship in Nurse and Midwife Education, December 1996 (see appendix 16). At this conference, a presentation of "A Mentor's Guide to Assessing Student Nurses" (appendix 1a) was given, followed by the distribution of a short questionnaire (appendix 17). A paper was later published in the conference report [468].

Conference questionnaire results.

Thirty questionnaires were distributed randomly within a group of approximately sixty participants. Participants were asked to complete the questionnaire and return it in the stamped addressed envelope supplied. Fourteen were returned giving a response rate of 47%, which is not unusually low for mailed questionnaires [2].

Respondents were lecturers, senior lecturers, senior nurses, ward managers, and mentors. Several were anonymous. Responses were generally positive or very positive. Some concern was expressed about the compatibility of the program with different types of computer. It was perceived as a useful supplement to traditional teaching, however it was considered too 'adult nursing' oriented and would have little appeal to others such as those from a 'mental health' environment. The content of the program was found to be relevant; the assessment process portrayed in the program was similar to that being used in most respondents' institutions. However, there was uncertainty as to whether the 'Concept Demonstration' part of the program was appropriately focused. One considered Benner's framework inappropriate; others had not heard of Bondy. The simulation was found to be realistic, effective, motivating and of high quality. The usefulness of the program for training mentors was perceived to be limited by the availability of computers and computer skills, especially in clinical areas. Overall, respondents were either positive or very positive about the program. The major concern was the sound and video quality. The questionnaire responses have been combined with evaluation data from other sources in section 8.4.5 (p135).

International conference presentation

'Nurse Education Tomorrow' International Conference, September 1998 [469].

A second prototype program (appendix 1b) and a paper outlining the strategy for its summative evaluation were presented, and warmly received. The strategy for evaluation was discussed: the development of an instrument utilising self-efficacy theory was considered innovative and appropriate.

Competitive 'Innovative Teaching' event

This was held at Glasgow Caledonian University, January 1998 and open to outside competitors. A joint presentation was given with the multimedia programmer who contributed to the production of 'A Mentor's Guide to Assessing Student Nurses'. A short paper was written and later included in the proceedings of the event. The program received a "Highly Commended" award, in recognition of its imaginative instructional design and evaluation strategy.

8.4.5 Formative evaluation results and discussion

Introduction

The results of the formative evaluation of the early prototype (appendix 1a) have been combined and summarised under the categories identified by Tessmer [453]. Recommendations arising from the evaluations follow. The next versions of the program in the appendix (1b and 1c) have incorporated the recommended changes.

Learning effectiveness

(a) Clarity of writing/narration

The level of language was acceptable for on screen reading. The educational psychologist thought that the introductory text was too formal, but the 'new mentor' group found it easy to read. Misunderstandings were few and mainly concerned ambiguities in the Q&A section of the 'simulation'. The purpose of this section was poorly explained and was not immediately obvious, causing some confusion. A misunderstanding was caused by ambiguity in the early title "Assessing Student Nurses": some at the 'Innovative Teaching' event initially thought that the program was to 'test' student nurses, rather than being a preparation for mentors. Some terminology was considered potentially puzzling to users, for example, one 'content expert' thought that the word 'taxonomy' was educational jargon.

(b) Sequencing of content

This was problematic for 'content experts' to varying degrees. A linear structured program was expected. The open structure of the program, with no predetermined

sequence, caused confusion and disorientation. It was also thought that 'freedom' to explore would lead to time wasting by mentors. It was advised that those parts of the program that were best tackled in a linear sequence, for example, data gathering in the 'Simulation', should give instructions to that effect. Both senior lecturers suggested that perhaps too much theory was presented at the *beginning of the program*, suggesting that they too expected a linear program, with a beginning, middle and end.

Others such as the 'new mentor' group found freedom to navigate the program neither surprising nor difficult. It was found that information was concise and easy to understand. Backwards and forwards navigation through text pages was thought necessary by some, and video controls needed to be made more apparent. For example, 'content expert 1' wanted to stop and review a section of video but did not know how to operate the controls.

The 'timed' pages of the introduction changed too quickly for some users who could not read the text before they moved on. The educational psychologist's advice was to allow users to be in control of this, as reading speed was bound to vary.

(c) Effectiveness of strategies.

The strategies to achieve the aim and objectives of the program (part 7.1, p 110) were considered appropriate by the 'content experts' and 'new mentor' group. It was unanimously agreed by content experts that the assessment simulation and model of the assessment process were acceptable representations. Using a variety of media and 'real' people were considered to be important motivating factors. The educational psychologist was captivated by these aspects. The 'new mentor' group found the experience of pitting themselves against the 'computer' to be motivating. Providing different 'frameworks' for judging the student's level of skill was found by 'content expert 3' to be a useful strategy.

Some areas of concern were identified. For example, one content expert found it unrealistic that the student's portfolio was available to the assessor to read as part of the simulation. However, on discussion it was considered appropriate, as similar information

would probably be provided by the student in discussion with their mentor. A feature of the early prototype was to limit the 'question and answer' selection to a maximum of three questions. As the computer contained the "correct" answer based on all the information in the case study, this was considered unfair by several participants, and was recommended to be removed.

(d) Reality of examples used in the program

The content experts considered that using student nurses filmed in real situations was better than using actors to stage nurse-patient interactions. The educational psychologist, however, believed that well scripted actors would have the advantage of being able to encapsulate issues in a shorter time frame, although this would be at the expense of fidelity.

It was also felt that the *hypothetical* nature of the 'question and answer' section of the simulation should be consistent with the case study within that part of the simulation. Others concurred with these generally positive findings; for example, the 'new mentor' group found the case studies to be realistic.

(e) Workplace performance

This was difficult to ascertain, although content experts and others knew that the program was intended for use by mentors in the workplace, and their comments were made in that context. Questions related to hardware compatibility were raised, as many hospital environments did not have Apple computers. Despite this, subjects who were practising nurses were positive about the benefits of having the program as a workplace-based resource.

(f) Amount of practice

None of the subjects 'exhausted' the program during the evaluation. From this, it was concluded that there was sufficient material for most users' needs. Both 'senior lecturers' asked if the intention was to incorporate more cases, perhaps recognising the limitations of having only four case studies in the simulation. The program has been designed to

incorporate additional case studies if the need developed. The mechanism of providing feedback in the simulation has been created specifically for this purpose. The limiting factor is the amount of data that could be contained in one CD-ROM.

(g) Quality of feedback

Generally, feedback was considered satisfactory, but with some reservation. The limited feedback in the 'self-assessment' sections concerned content experts 1 and 2. It was suggested that users ought to be told whether they were correct or not, and given reasons. The feedback statement following the entry of the user's assessment in the simulation was found helpful by 'content experts 3 and 4'. However, 'content experts 1 and 2'would have liked further elaboration, as simply being told how many domains were correct was not enough for them. This was supported by the 'new mentor' group, and was similar to the initial reactions of the educational psychologist, computing consultant, and by several participants at the 'innovative teaching event'. However, following discussion, it was agreed that there was a strong motivational element in *not* simply being given a full answer, prompting the user to try and work out the correct level of skill for themselves by reviewing the evidence. Indeed this is precisely what happened in the 'new mentor group', who were determined to do better.

Several, including two content experts, were of the view that it would add credibility to be informed that the 'correct' scores were based upon the judgement of experienced mentors.

(h) Quality of learner interactions.

This varied within the program. Through observing those who used the program and in analysing their comments, the program was capable of capturing and sustaining the user's attention. During one-to-one and small group evaluations, none of the subjects seemed bored or wanted to finish quickly. Users were challenged and encouraged to think carefully, especially evident during the 'Simulation'. These observations would indicate that the program can facilitate interaction at the higher levels, as outlined in appendix 7.

(i) Coherence of graphics and layout

The experts generally found the program to be visually pleasing and the graphics to be of high quality. Some aspects were criticised, for example, some screens seemed crowded. The numbers, 1-5 in the assessment grid was found to be ambiguous. It was not clear whether 1 was the lowest or highest level. Further explanation was felt necessary. Colour coding on the 'model' diagram was inconsistent, and meaningless. These were minor flaws.

(j) Suitability of icons

These were found to be representative and unambiguous, with several exceptions.

The icon representing 'Orientation' in the Model of the Assessment Process was thought to be clickable at each point, N, S, E and W. Similarly, the seven thumbnail photographs that constituted the 'Domains' icon, were each thought to be 'clickable'. However, these problems were encountered by only one individual. Thumbnail photographs in the 'question and answer' section were confusing, as the staff and students were different to those in the case study. Alternative ways of navigating were discussed with the educational psychologist who liked the interface but thought it could possibly be improved by the use of pull-down menus, similar to those in Windows ™ applications. However, there was no support for this in any of the comments made by others, who experienced little difficulty.

Summary and conclusion: the educational effectiveness of the program

Despite some reservations, the program has been reasonably effective in providing the context for learning. In places, such as the 'simulation' it requires cognitive engagement at a deep level, decision-making and prioritising. It has been found to be motivating and engaging.

Content quality

(a) Content accuracy

The program was found to have a small number of elements open to misinterpretation, however, there were no major inaccuracies. Debate surrounded the 'correct' score, but subjects were content that these had been decided by an expert panel.

(b) Content currency

The content for prototype (appendix 1a) was created during 1996. Certain aspects have since been in need of updating, for example, Glasgow College of Nursing no longer exists, and some of the documents, policies and icons within the program required to be changed. Similarly, the role of the 'link teacher' continued to develop, for example. These issues were known to the researcher, and were not of concern. One of the senior lecturers advised that 'in – text citations' should be removed because of their tendency to date quickly. It was encouraging that the 'questionnaire respondents' and the participants at the International Conference identified similarities between the process of assessment in their own institutions, although none were identical.

(c) Content 'completeness'

The program was perceived to have several omissions. The video content made it possible to observe limited aspects of a student's performance. However, as this was likely to be the case in real situations, arguably this was not an omission, but reflected reality. One content expert felt that the student's self-assessment would have been useful in each case study. Another felt that a printout of references would have been beneficial for users. An alternative framework to Benner's would be an enhancement (although no alternative was suggested). Lack of detail to explain each of Benner's Domains was criticised by the educational psychologist, who suggested that practical examples could highlight the distinction between domains. These were worthwhile suggestions that were considered further.

(d) Content 'superfluousness'

The instructions before each main part were found by 'content expert 3' to be repetitious and irritating. A suggestion was to allow the user to read the introduction only once on entering a section. Thereafter, a help facility could produce these instructions, but only if required.

Summary and conclusions: the content quality of the program

The program was found to have acceptable content, neither too little nor too much, with a small number of recommended additions. It was also considered an accurate representation of the issues surrounding the assessment of student nurses. In terms of currency, it was not surprising to find that aspects of content were out of date shortly after being included in the program, given the enormity and rate of change in nurse education, both locally and nationally. A number of changes were therefore recommended.

Implementability

(a) Teacher ease of use

The process of installing folders onto the hard disc was familiar and no difficulties were envisaged in this process, except by the 'naïve users' who were anxious at the prospect. Program content was familiar to lecturers. The 'content experts', considered that the program would be of benefit in preparing mentors if enough appropriately equipped computers were available.

(b) Learner ease of use

There were difficulties encountered by some. The model diagram was designed to be an 'advance planner' but was only partially successful in this regard. At the time of these evaluations (1996 – 1997), the Internet was not widely used, and the concept of hypertext links was still new. None of the content experts, nor mentor group knew to click on a highlighted word, even although the instruction to do so appeared at the foot of the appropriate page. In addition, the 'learning outcomes' for the program were not immediately apparent. One 'content expert' frequently compared the program to others

with which he was familiar. These had been linear and behaviourist in structure and he could not understand, for example, why this program did not state explicit 'objectives'. Others (content experts 1 and 3) seemed to appreciate that learning outcomes for individuals depended upon their previous knowledge and required to be formulated by the user and thought this constructivist approach was facilitated by the chosen program design.

Assessment data entry into a grid came as a surprise to some users as there was insufficient information to this effect earlier in the simulation. Several buttons were ambiguous, for example 'go back' and 'quit'. Some aspects of the 'model of the assessment process' diagram was also found ambiguous: the red dots in the line diagram were thought by 'content expert 3' to be clickable links. Others evaluating the program had only minor difficulty with navigation, and soon became familiar with it.

(c) User willingness to use

This was not possible to determine with everyone during formative evaluation: however, 'questionnaire respondents' confirmed a willingness to use such a program, as did the group of 'new mentors' who were particularly enthused by it.

(d) Fit to learning environment

There was concern expressed in relation to incompatible or inadequately powered hardware, especially among those 'questionnaire respondents' based in clinical rather than educational environments. The 'new mentor' group thought it would be useful to have the program either in the ward or accessible elsewhere in their hospital.

(e) Orientation and support requirements

There have been questions raised regarding compatibility, navigation, and learning outcomes of the program. Both senior lecturers suggested an accompanying instructional booklet with references.

(f) Administrative acceptance

This was difficult to judge. However, one of the content experts was responsible for the mentorship course and its development, and commented that the program could be used in the university when it was available. Questionnaire respondents, some of whom were clinical managers, were generally keen to have the program available for immediate use. The overall impression was that if it was technically possible, then nurses in both education and clinical environments would find the program beneficial.

Summary and conclusion: program implementability

There have been some minor problems, for example, ambiguous buttons and hypertext links. The absence of information on possible learning outcomes could be remedied, as could better instructions be included for data entry in the simulation. However, the major concern of a number of subjects was the availability of suitable hardware to allow the program to be used in typical clinical environments. Most hospital computer systems have been based on Intel or IBM, but not Apple Macintosh. There was therefore a need to overcome this problem by exploring the feasibility of creating a version of "Assessing Student Nurses" that was cross-platform compatible.

Learner interest/motivation

(a) Interest in content

The program was novel, stimulating, and capable of motivating users. As the variety of media varied within different parts of the program, so too did the level of concentration, interest, and cognitive interaction on the part of the user. The senior lecturers considered the introduction academic, too long and boring. It was suggested that the user be given enough information to allow an informed choice of route through the program.

(b) Level of learner challenge

This was difficult to ascertain as only certain aspects of the program were sampled by any individual or small group. 'Content expert 2' suggested a hierarchy of difficulty would be beneficial in the case studies and self-assessment sections. In reality, some users found certain case studies more difficult than others did, for no obvious reason.

Some content experts scored highly in the simulations, others scored mid-range: it could be concluded that the level was sufficiently challenging, especially for new mentors.

(c) Willingness to learn more/explore

Each of the content experts was willing to explore and browse the program, even after two hours had elapsed. Mentors, too, were eager to investigate different aspects of the program, and concentration on content was sustained for long periods.

(d) User willingness to use

One content expert was initially reluctant to have her judgement 'tested' by the computer. However, this disappeared on receiving positive feedback, and finding that 'the computer' was largely in agreement with her. The 'new mentor' group were not as defensive, keen to enter their judgement and vehemently questioned the validity of the correct version against which they were compared. One could reasonably conclude that there is sufficient of interest in the program to ensure its use.

(e) Perceived value of learning

Overall, this program was perceived as a valuable tool in the facilitation of learning. This was a widely expressed view, with no negative comments in this area.

(f) Time spent learning

In common with other multimedia programs, where discovery and browsing are features, time may not be used as effectively as in 'programmed learning' types of programs. Hence, efficiency of learning may be affected. The case for designing an openly structured programme therefore lies with quality issues. Learning has been argued to be of higher quality because incidental learning occurs through browsing, and because the user can control route, sequence and level of learning, carefully tailoring it to their needs. 'Content expert 2' expressed the view, on several occasions, that mentors would need to be directed through the program in order that the best use would be made of their limited time.

With discussion within the group of mentors, one case study took 45 minutes to complete.

Summary and conclusion: learner interest/motivation

Mentors and questionnaire respondents were overwhelmingly positive in their attitude towards the program. The educational psychologist found the program enjoyable. Each of the four content experts was so engrossed that they wanted to continue with the program even after completing the evaluation exercise. There would seem to be sufficient challenge within the program to provide interest and keep learners 'on task'. This was evident throughout the evaluation process

Technical quality

(a) Visual quality

The quality of video, although poor in places, was not universally condemned, nor did it detract from users' enjoyment of the program. The fact that a computer was capable of showing video at all seemed to compensate for the fact that there was poor sound, and loss of lip synchrony in places. Otherwise, the visual appearance was well received.

(b) Aural quality

This was variable, from acceptable in some places (the comments of 'actors' in the 'judging performance' part), to poor in others (case study videos).

(c) Effective use of media

The variety of media used in the program appears to have achieved an appropriate balance. The 'content experts' thought that there was a good mix, as did the 'new mentor' group. The computing consultant was impressed by the range of media and thought this would maintain user interest. Integration was an issue in places, for example, the absence of a link between the photographs and text in the 'question and answer' section of the 'Simulation'. Both senior lecturers thought that there should be more sound, and less text, whereas the educational psychologist and the computing consultant advised against audio being widely used because users differ in how well they attend to verbal instruction. This

contrasts with text that can be read for as long as it takes for it to be understood. The educational psychologist thought that more media could be useful; for example, a photograph of Benner could add interest to an otherwise monotonous text screen.

Summary and conclusion: technical quality

The range of media included in the program served its purpose in maintaining user interest. In addition, different approaches were found to enhance the learning experience, especially in the simulation part of the program. Unfortunately, the memory requirements of video and audio have made compromise necessary. Monochrome video was not successful, and some audio required to be re-digitised. These were perhaps minor problems in a program that has been shown to have a high degree of educational value overall.

8.4.6 Concluding comments: formative evaluation

There were many aspects of the prototype version of 'Assessing Student Nurses' (appendix 1a) that required attention, as detailed in this summary of formative evaluations. However, the main structure of the program was sound and remained intact, as have the essential elements of the interface, which were confirmed as being effective. The formative evaluations have thus been very useful in distinguishing areas of strength from areas of weakness, and have given clear indication as to how to complete the program to a high standard.

8.4.7 Recommendations (grouped in appendix 18.)

Recommendations followed from the analysis of the collated results of all formative evaluations. This was used as a work plan to produce a program of much higher standard than the original. The resultant programs are included as appendix 1b and 1c.

Chapter 9: Materials and Methods

9.0 Summary

This chapter outlines the development of data collection tools, and research methods. Firstly, the data collection tool for the RCT was required to collect data pre and post educational intervention (program or traditional). The approach, based on the findings of the literature review, comprised measuring anxiety, attitude, and self-efficacy. This was acceptable to 15 experienced lecturers through focused interviews, providing evidence of concurrent and face validity. Anxiety was measurable using Spielberger's State Anxiety Inventory; attitude and self-efficacy items were developed from the literature review. An adapted q-sort was used to assess these items for relevance and precision. The final 65-item instrument demonstrated internal consistency, test re-test reliability and was acceptable to prospective mentors in trials providing further evidence of face validity. A second instrument was designed to guide focussed telephone interviews in a selected sample of subjects post-RCT.

Issues in the organisation of the RCT are then discussed, including the sampling strategy, randomisation, bias, and ethics.

9.1 Development of data collection tools

9.1.1 Introduction

Two instruments were required. Firstly, to achieve Aim 2, objective 7.2.1, a research instrument was required to collect quantitative data in the RCT. Rigour was required during development to establish acceptable validity and reliability. Secondly, a protocol for focussed telephone interviews was required in order to achieve objective 7.2.4.

9.1.2 Development of the RCT data collection instrument

Based on the literature review, assumptions were made regarding the relationship between anxiety, attitude, knowledge and self-efficacy and the quality of student nurses' assessments. As discussed in section 5.5.4 (p68), some anxiety can enhance performance,

being motivational in nature. However, excessive anxiety is detrimental, inversely correlating with performance. This contrasts with the positive correlation that exists between performance and mentors' knowledge of the assessment process (5.5.6, p71), their attitudes (5.5.2, p66) and self-efficacy in relation to their assessor role (5.6.2, p73). Although these assumptions arose from the literature, it was considered appropriate to subject these to further scrutiny. The objective was therefore to find out whether lecturers generally believed that knowledgeable mentors, not over- anxious about assessing, with positive attitudes and high self-efficacy, assessed students in a way that accurately reflected their abilities.

The validity of literature review findings

The views of experienced nurse lecturers provided an element of concurrent validity to the findings of the literature review. Focussed interviews sampled the views of lecturers, rather than mentors, on the grounds that those selected prepared mentors, linked with clinical areas, supported students on placement and dealt with the consequences of their assessment. Whereas mentors would not admit, for example, to anxiety or attitudes affecting their judgement, lecturers could be more objective. Their extensive experience provided an insight into factors that indicated mentors' assessment ability. The original intention was to interview twenty lecturers, however, after the fifteenth it became apparent that little more was likely to be achieved and the process was terminated.

Sampling and ethical considerations

A purposive sample of lecturers from one university was selected, to represent each of the branches within a nursing programme. Each lecturer had a minimum of six years experience in the support of students on placement. An explanation of the purpose of the interview and assurances of confidentiality were given. Consent was obtained.

Interviews

A topic guide, as detailed in appendix 19, was designed to provide a narrow focus making it unnecessary to record the interviews, as responses were noted as they occurred. Interviews lasted less than 30 minutes. Focusing participants on the topic area was

difficult at times; extrinsic factors such as shortages of staff and workload pressures were frequently raised. Data was readily categorised under the topic questions. This information was collated and summarised, as follows:

(a) Topic area: anxiety

Subjects found that most mentors worried about their role, and that this affected their ability to assess students, especially when the mentor was new or inexperienced. However, some subjects had reservations about anxiety being a useful indicator of assessment ability: some mentors were naturally anxious. The optimum level of anxiety for an individual was considered difficult to gauge, as it was appreciated that a certain amount was required. One subject considered that the term "confidence" was more appropriate than "anxiety", and felt that this would have a much clearer relationship with assessment ability.

One subject initially claimed not to have experienced much anxiety in mentors, but on reflection, concluded that some behaviours could have been evidence of anxiety, for example, one mentor always asked questions, some of which seemed irrelevant. Other lecturers qualified their views by commenting that mentors who appeared to have the ability to assess did not appear to use it: several examples were given of mentors passing students who should have failed. Anxiety in relation to the consequences of failing a student possibly contributed to this, but factors such as motivation were also considered important.

(b) Topic area: mentors' attitudes

Mentors' attitudes were found to vary considerably. Positive attitudes were associated with capable mentors and negative attitudes characterised mentors who had difficulty with their role. Subjects were not clear whether this was a causal relationship. Explanations for mentors' negative attitudes were proposed, such as being conscripted into mentorship by management, or having had difficult experiences with past students. Being expected to assess students also caused negative attitudes, as some held the view that this was the role of an academic, not a practising nurse. Positive attitudes seemed to

develop through satisfaction from the mentor-student relationship, where teaching and the facilitation of learning were particularly rewarding. There was agreement that mentors' attitudes affected how they performed their role, and how students' assessments were performed, for example spending time with students, reading and understanding course documentation, and paying attention to detail.

(c) Topic area: knowledge and understanding of the assessment process.

The subjects considered that knowledge and understanding had an obvious connection with the ability to effectively assess students. Assessment tools and documentation were often complicated, subject to change and not always self-explanatory. It was generally believed that having a working knowledge of the assessment process was essential for undertaking the mentoring role. However, not knowing when or how to assess students or how to complete documentation were common findings and often required a clinical visit by the lecturer to explain the process. Stereotypical, inappropriate or inadequate written comments accompanying the student's assessment were thought to reflect a lack of understanding on the mentor's part. A commonly expressed view by the subjects was that knowing the principles of assessment and how to approach it from a theoretical perspective, was no guarantee that a fair and accurate assessment would result.

(d) Topic area: self-efficacy and other intrinsic influences on the mentor

Miscellaneous factors such as motivation, pleasing strategies, conformity to others' expectations, being in control, and self confidence were thought to have some effect on mentors' assessments. Many respondents were of the opinion that anxiety, knowledge and attitudes provided a comprehensive account of the influences on mentors. In addition, the 'self-efficacy' of the mentor was strongly believed by some to be a very important factor. Some equated this with self-confidence and self-belief and had a crucial role in determining how well mentors performed. However, several respondents were unsure of the concept and were unable to comment.

Discussion of findings

The findings from these interviews reinforced the findings of the literature review: mentors' anxieties, knowledge, attitudes and self-efficacy can affect the quality of students' assessments. It was not surprising that a proportion of lecturers were unsure of self-efficacy theory, as it has been rarely reported in mainstream nursing literature. However, those that knew of it, considered it to be very important. As these factors are measurable, see sections 5.5 (p66) and 5.6 (p72), it should be possible to construct a research instrument to indicate how well a mentor is likely to assess students. Lecturers tended to qualify their comments in a way supportive of mentors, demonstrating an insight into the difficulties and conflicts associated with the role.

Item identification and scale construction

It was important to identify any research instruments that were available or could be readily developed, in order to benefit from the results of validity and reliability testing. Some scales such as self-efficacy require to be context specific, whereas others such as anxiety can be more generalised. Streiner suggested that when constructing a scale, identifying items is a process of gathering from a variety of sources and then weeding out those that do not meet the required criteria [299]. The items were therefore generated from existing scales, research findings and theory. Some points raised by lecturers during interview, for example, mentors' negative attitudes caused by having an increased workload when mentoring a student, were useful as items to include in the questionnaire.

(a) Anxiety item identification and instrument design

A scale that has been widely accepted as valid and reliable was selected (5.5.5, p71), the 'State Anxiety Inventory', one half of the STAI by Spielberger et al [298]. Licence was obtained from Mindgarden ™ [300]. This was used with no changes to content, but presentation was adjusted to be consistent with the remainder of the questionnaire: see appendix 20, items 9-28.

(b) Attitude item identification and instrument design

Attitude items related to being the assessor of students' performance, and associated attitudes towards students and nurse education were identified from literature and the focused interviews with lecturers. It has been possible to phrase these attitude items either positively or negatively, in order to enhance the reliability of the instrument. A 5-point Likert scale was constructed, based on the advice of Oppenheim [272]. These are items 29-50 in appendix 20.

(c) Knowledge item identification and instrument design

This presented fewest problems in instrument development. Knowledge items were presented in the form of questions with a choice of possible answers, only one of which was correct. This was selected by ticking a box, similar to a multiple-choice test. Items were derived from the literature review and were carefully worded to avoid ambiguity.

(d) Self-efficacy item identification and instrument design considerations

It was decided to measure 'strength' of self-efficacy, using a confidence continuum. Items were derived from the 'process of assessment' as outlined in the literature review, section 4.8 (p53), and appendix 5. Subjects could indicate how confident they were in relation to the performance of each item, on a confidence continuum. This is shown in items 51-73 of appendix 20. Due to the shortcomings of many self-efficacy instruments, as explained in section 5.6.8 (p79), tests for validity and reliability were required.

Sorting the items

Having identified items for possible inclusion in a questionnaire, a q-sort procedure [2] was adapted to determine how well each item represented its 'domain'. Anxiety items were not included in this process, as the STAI had already been validated. Twenty attitude, twenty self-efficacy, and nine knowledge items therefore comprised the pool of items. Each item was printed onto a card, shuffled then numbered randomly from one to

forty nine. A purposive sample of six lecturers was drawn from the pool of fifteen previously interviewed.

Subjects were asked to identify

(a) Whether the item indicated self-efficacy, attitude, or knowledge

(b) How well it represented these, by placing on a continuum from 'not a good measure' to 'a very good measure'.

Results identified difficulty placing eight items, some of which were removed; others re-worded to avoid ambiguity. Several items that were constructed as 'self-efficacy' items were mistakenly classed as 'knowledge' items, and vice-versa. This was possibly because having knowledge can be a major element to being confident in performing a task (173).

Face validity of the instrument

This was ascertained through the scrutiny of lecturers and mentors.

Lecturers

Items were adjusted to remove ambiguity and weaker items were winnowed out. Six copies of a draft version of the compiled 'assessment potential' instrument were distributed to lecturers who took part in the initial interviews, for comment. This confirmed that the instrument was capable of measuring what it was intended to measure, providing face validity. Additional grammatical, aesthetic and presentational comments were helpful.

Several comments expressed displeasure at the structure of the "knowledge and understanding" questions. It was perceived as resembling an 'exam' rather than a questionnaire and it was felt this would not be well received by subjects. Given the difficulties experienced in the Q-sort, the usefulness of knowledge items was reconsidered. As the possession of self-efficacy in relation to the performance of a skill reflected the possession of sufficient knowledge of what was required, then the separate measurement of knowledge could be argued to be superfluous. As this section of the questionnaire could have reduced the compliance of subjects, then it was removed.

Several self-efficacy items were created from the knowledge questions, for example the knowledge item:

"*Before a clinical assessment can be said to be valid, which of the following criteria need to be satisfied: select from the following list of alternatives...*

This became the self-efficacy item:

"*As of now, how confident are you that you are able to state the criteria that determine whether or not an assessment is valid*"

The response to this statement on a Likert scale from 'very confident' to 'no confidence whatsoever' would, according to Bandura's [173] theory of self-efficacy, correlate with an individual's ability to state these criteria, which in turn would be dependent upon their knowledge of them. In this way, the acquisition of knowledge stated in objective 7.2.1, was not assessed separately as an outcome, but was implicit within the carefully constructed self-efficacy items 58, 65, 67, 68, 71 and 73 of Appendix 20. This had the advantage of indicating the subject's knowledge of the assessment process and increasing the likelihood of compliance with data collection, while minimising the size and complexity of the questionnaire.

An introductory section requesting biographical details from participants were added, as shown in appendix 20, items 1-8: appendix 20. This information was intended to provide data to achieve the objective 7.2.2, and to test the null hypotheses 7.2.2.1 and 7.2.2.2.

Prospective Mentors

The aim was to discover how usable the questionnaire was when used with a group of subjects. Eleven prospective mentors attending a half study day, the subject of which was clinical assessment, consented to completing the form before, then again after the session. It was found that the maximum time to complete was 13 minutes before the session, and 10 minutes at the end. This was acceptable for the Randomised Controlled Trial. There were no difficulties or ambiguity reported. A superficial analysis of the questionnaires

demonstrated a noticeable before and after 'improvement' in many items. This finding was important in helping to judge the sample size in the Randomised Controlled Trial (9.2.3, p158). Informal discussion with this group concluded that the instrument was acceptable, contributing to its face validity.

Establishing test-retest reliability and internal consistency

This was required to achieve objectives 7.2.1.3 and 7.2.1.4 of Aim 2, section 7.2 (p111). A well-constructed scale should be stable over time, assuming nothing changes in the interim [299]. It was assumed that anxiety, attitude, and self-efficacy in relation to observational assessment were stable attributes that did not vary without cause. The consistency of scores on two separate occasions in a sufficiently large group of people could therefore be taken as an indicator of the stability of the measurement instrument and reflect on its construction. This was the rationale for determining test re-test reliability. Streiner suggested that the re-test should not occur within two weeks of the initial test but the time interval should not be so long that the variables being measured had changed [299]. The data collected for test re-test reliability was also used to test the internal consistency of the instrument (alpha).

Testing the questionnaire

Prospective mentors in four groups of 12-18 completed the instrument at the end of half-study days. The re-test was posted to participants two weeks later with a cover note (appendix 21), and a stamped addressed envelope.

Results

Fifty-eight mentors completed the initial questionnaire, forty-three postal responses were returned providing forty-three pairs for comparison. The response rate was 74%, better than could normally be expected from a postal questionnaire [2]. This was undoubtedly enhanced by follow-up telephone calls.

Questionnaire scoring

The score sheet accompanying the SAI [298] guided the scoring of anxiety items. Attitude items were scored 1-5, '1' being indicative of negative attitude. Hand scoring templates for both anxiety and attitude were produced, to make scoring easier and more reliable. Self-efficacy was scored from 1-5, '1' being low self-efficacy. Data was entered into the Statistical Package for the Social Sciences (SPSS™); the results of analysis summarised in table 14.

Table 14: Results of test-retest reliability and internal consistency

	TEST RE-TEST CORRELATION	ALPHA
Anxiety	0.73	0.93
Attitude	0.70	0.74
Self-efficacy	0.82	0.94

Results

The instrument demonstrated acceptable internal consistency, with correlation co-efficients greater than 0.70 [2, 299]. No single item, when omitted, significantly increased the reliability of the sub-scales when re-tested.

Discussion.

Self-efficacy has a closer test re-test correlation than either anxiety or attitude scales: this may be a function of the design of the five point scale, where items are phrased in one direction, from low to high. This may have made errors of central tendency and stereotypical entries more likely. It was interesting to find that the test-retest correlation results for anxiety were consistent with Spielberger's normative data for college students (0.73 to 0.86 [298], p32). Spielberger[298] asserted that given the nature of anxiety states, measures of internal consistency such as the alpha coefficient provide a more meaningful measure of reliability. This assertion is supported by the data in table 14.

Conclusion

The questionnaire demonstrated acceptable test re-test reliability and internal consistency. In addition to face and concurrent validity, this demonstrated that the instrument was stable and could be used reliably.

9.1.3 Protocol for telephone interviews

It was anticipated that the quantitative data from the Randomised Controlled Trial would benefit from elaboration. For this reason, it was planned to interview a purposive sample of subjects to obtain qualitative data focused on areas of interest. The exact criteria for selection of subjects was dependent upon the results of data analysis at the end of the trial. Once this was clear, a protocol was produced (appendix 22) to guide the interview. SPSS™ was programmed to select subjects to meet the required criteria.

9.2 The planning and organisation of the RCT

9.2.1 The population

"A Mentors Guide to Assessing Student Nurses" was produced for mentors who assessed students. As this combination of roles was not universal within the UK, the program was mainly applicable to mentors in Scotland. The lack of accurate data on mentor numbers in Scotland, as discussed in section 4.4 (p44), led to a recommendation by Watson and Harris [13] that institutions ought to maintain a live database of mentors, with information on their location and preparation. This was pre-empted in Glasgow Caledonian University by the creation of such a database in 1998, for internal administrative purposes. Each record comprised a trained nurse working in the Greater Glasgow Health Board Area, hospital or community, private or public sector, who had applied to attend a programme of study in order to become a mentor. Ten months following its creation, the database held 759 records, indicating a high level of activity. Although the outcomes and duration of mentor preparatory courses varied throughout Scotland [13] applicants were presumed to have similar characteristics. Generally, these were registered nurses with a minimum of 1 year's experience. There was nothing to indicate that applicants to Glasgow Caledonian University's mentorship preparation

programme were anything other than characteristic of the population of prospective mentors in Scotland.

9.2.2 Sampling and ethical considerations

Eligibility criteria were that prospective subjects required mentorship preparation as part of their job, had been qualified for more than 1 year and had not previously undertaken a mentorship course.

From an organisational perspective, it would have been ideal had experimental and control groups been randomly formed and consent obtained from subjects just before each session. This could not happen because consent for randomisation itself was required, not merely for the different treatments applied to experimental and control groups [458]. This made it mandatory that information was given prior to randomisation and was the basis for ethics committee approval (appendix 23). The trial was publicised and prospective mentors invited to apply (appendix 24). For those that did, consent for randomisation was implied. However, this process may have weakened how well the volunteer sample represented the population of mentors at large. To minimise the introduction of bias, these volunteers were allocated to either experimental or control groups using a table of random numbers. Some negotiation over dates was allowed. As the control group was the traditional teaching method, this was a smaller number of slightly larger groups composed of those applying to the research, but infiltrated by those applying through the usual university mechanisms, or by just turning up on the day because, for example, their ward was quiet.

9.2.3 Sample size

The effects of teaching the observational assessment of performance on mentors' anxiety, attitude and self-efficacy was unknown, therefore a power calculation to determine the required sample size could not be performed. An estimate of the sample size was made based on the expected effect and whether this was likely to be small, medium, or large [470]. Polit and Hungler advised that a medium effect should only be estimated when it is easily observable, otherwise a small effect should be assumed [2]. The process of

developing the research instrument provided an opportunity for gathering data from eleven prospective mentors, as explained in section 9.1.2 (p147). A detectable 'before and after' difference was found allowing the assumption that a medium effect could be expected. Jaccard and Becker advised that for an alpha value of 0.05 and power of 0.80 the total sample size for a study with expected medium effects would be 80[471]. This estimate took no account of dropouts from the trial: Hillier[472] advised that a study should continue until the required numbers had been achieved. This size of sample was estimated to detect a statistically significant change in pre and post score based on traditional teaching methods. However, as the difference between multimedia and traditional teaching methods remained unknown, this sample size may not have been sufficient to allow any differences to become statistically significant. A full pilot study to compare the before and after effects of using the program could have provided this information, however time and resource constraints did not allow this. A larger sample size target of at least 120 subjects was therefore set.

9.2.4 Avoiding sample contamination

The production of materials for inclusion in the program, it's formative evaluation, and data collection tool development was confined to one of six NHS Trusts within Glasgow. Individuals from this trust were not included in the Randomised Controlled Trial. Other activities in formative evaluation, such as the seminar presentations and conference presentations did not include anyone in their audiences who subsequently became a subject in the RCT. Despite these precautions, contamination could have occurred through subjects discussing their experiences of the trial with prospective subjects. This was difficult to detect or to avoid as the trial was staggered over six months.

9.2.5 Mentor preparation course

The mentorship preparation course at Glasgow Caledonian University was vocationally oriented, consisting of a combination of half and full study days. Attendance for the equivalent of three days resulted in the award of a certificate, conferring the status 'mentor' on the individual. This mentor preparation used teaching facilities within each of the Health Board's six 'zones' to increase the ease of attendance for participants, who

came from both private and public sectors. The University produced timetables, and posted these to clinical areas. Prospective mentors, with permission from their managers, decided which of the sessions to attend.

9.2.6 Control group treatment

Teaching in each zone was by different teams of lecturers, organised to reflect local conditions. The topic "*assessment of clinical practice*" was a regular feature and a period from January 1999 to June 1999 was identified when this topic arose in the programmes of each zone. Seven dates were identified, to which control subjects could be allocated once randomisation occurred.

9.2.7 Demography of experimental and control groups

The presence of sampling errors could be identified by using the chi-square test to compare responses of study groups to questionnaire items 1-8, appendix 20, for significant differences. In addition, demographic data on 759 mentors held in the Glasgow Caledonian University database was available for comparison with study group data. The chi-square test would again be appropriate for statistical analysis, the results indicating how representative the study groups were of the population of mentors at large. Further comparison of experimental and control groups was undertaken by analysing baseline scores for anxiety, attitude, and self-efficacy, using the Mann-Whitney U test, to detect differences between groups. The hypotheses in relation to comparing experimental and control groups are stated in Aim 2, objective 7.2.2.

9.3 The RCT

9.3.1 Introduction

Ideal conditions for this research would be to treat subjects identically, the only difference being teaching method: traditional or multimedia. Close control of variables such as numbers in a group, time of day, setting, and teacher were desirable but not practicable due to variation between zones. Flexibility was thus required and permissible within a pragmatic randomised controlled trial, the rationale for which has been explained in section 6.8.9 (p107). Controls were in place in the following areas:

- Teacher introduction and purpose of the research
- Rapport development prior to administering the data collection instrument
- Duration of teaching session
- Conduct of the researcher and assistance provided in the experimental group
- Learning Outcomes.

9.3.2 The different treatments of experimental and study groups

Each experimental group session was held from 1.30-4.30. The computer lab was equipped with desktop computers, headphones and copies of the data collection instrument. Rapport was established with students before administering the questionnaire because some of the items in the SAI enquire about negative characteristics, for example, feeling tense or frightened. Some people would not have admitted to this, considering it a weakness.

The session started with a welcome and introductions. The purpose of the research was explained, and verbal consent gained. Baseline data was collected using the questionnaire. Subjects were given written instructions on starting the program as shown in appendix 25, however, the researcher gave initial help to those in difficulty. Once subjects accessed the 'guided tour', no further help was given. It was noted that there was frequent interaction between subjects, and several collaborated in tackling the case studies. After 2 hours, subjects were asked to finish what they were doing and complete a second data collection instrument. It was possible to form pairs of pre and post questionnaires by matching the handwriting of those subjects who declined to supply their names, which regularly happened. Spontaneous comments were usually positive and enthusiastic: the groups seemed happy with their session. Questionnaires were stored securely.

For the control groups, the location and timing of each session varied. These were normally in the afternoon in a small hospital teaching suite. The session began by the researcher explaining the purposes of the research, and obtaining verbal consent. Usually it transpired that several had no advance knowledge of the research, but willingly

participated. This was an expected result of the process of randomisation (9.2.2, p158). The questionnaire was completed at the beginning and at the end of the two-hour session, as in the experimental group. In between, traditional teaching was delivered by a university lecturer. Although the exact method varied, teaching was a modified lecture enhanced by discussion of assessment documentation and the role of the mentor in relation to assessment. Vignettes were sometimes used as a focus, and copies of assessment documents available for analysis and discussion. Despite slight variation in teaching method between lecturers, the learning outcomes of each session were the same. The program "A Mentors Guide to Assessing Student Nurses" shared these. Details of the teaching of 'Assessment' to the controls are given within Appendix 26. At the end of each session, subjects completed a second questionnaire, which was then paired with the first, coded and stored securely.

9.3.3 Focused telephone interviews

A follow up to the RCT addressed areas of interest arising from results in order to elucidate possible issues worthy of discussion. Short, focused telephone interviews were organised for a purposive sample of approximately 10% (n=12). The object was to explore possible reasons to account for differences in effect that the trial had on some individuals in both the control and experimental groups, as stated in section 7.2 (p111), aim 2, objective 7.2.4. This method is known as 'typical case sampling' [473].

Analysis of RCT data in both experimental and control groups identified individuals at the extremes: those that improved markedly, and those whose scores deteriorated in each of the sub-scales within the questionnaire. Once criteria were identified from the RCT results, SPSS™ was programmed to select cases for interviewing. The details of this follow the results of the RCT.

Chapter 10: Results

10.1 Summary

This chapter provides the results of analyses of data from the following sources: the RCT, data held on mentors in an electronic database; normative data from studies with Spielberger's STAI; and from telephone interviews. SPSS™ was used for statistical analysis.

Mentors recruited to the study were representative of Scotland in terms of their place of employment, gender, and clinical grade. Years since qualification, experience of assessment, educational background, professional qualifications, and frequency of exposure to students in the workplace, scores for anxiety, attitude and self-efficacy, were all similar between computer and control intervention groups at baseline. Finally, the anxiety profile of subjects was similar to that of a comparable sub-group.

Both the experimental and control groups showed statistically significant improvements in anxiety, attitude, and self-efficacy following educational intervention. Differences between study groups were not significant for anxiety and self-efficacy. However, there was a statistically significant improvement in pre and post measures of subjects' attitude in the multimedia teaching group compared with the traditional teaching group.

Telephone interviews conducted following the trial focused on those whose scores either greatly improved or deteriorated, in both the experimental and control groups. Categorised comments indicated little change from the immediate post-test data. The interviews indicated that cognitive conflict may have had a confounding effect: the disconcerting effect of being challenged on long-held beliefs increased anxiety in some subjects, negating the reduction in anxiety that might otherwise have been the case. This differed from the heightened anxiety that was experienced in some subjects in the control group, who expressed frustration and disappointment. The main benefit of the computer teaching session was directly related to the impact that the content of the program had on the user's ability to assess students.

10.2 Randomised Controlled Trial results

10.2.1 Comparison of control, experimental and mentor database groups.

The Glasgow Caledonian University mentor database contained three categories of data identical to items in the 'assessment potential' questionnaire used during the RCT: workplace, gender and clinical grade. This was used to test the hypotheses 7.2.2.1.

Table 15: Constitution of groups by workplace

	SPECIALISED UNIT	HOSPITAL WARD	CLINIC OR DAY HOSPITAL	NURSING HOME OR CARE OF THE ELDERLY HOSPITAL	GP'S SURGERY OR HEALTH CENTRE
Mentor Database, n=731	117 (16.0%)	375 (51.3%)	23 (3.0%)	55 (7.5%)	161 (22.0%)
Experimental Group n=62	9 (14.5%)	38 (61.0%)	2 (3.0%)	4 (6.4%)	9 (14.5%)
Control Group n=68	17 (25.0%)	31(45.0%)	1 (1.4%)	10 (14.7%)	9 (13.2%)

NB. 28 Records on mentor database were incomplete.

Table 16: Constitution of groups by gender

	MALE	FEMALE
Mentor Database n=709	67 (9.4%)	642 (90.6%)
Experimental Group n=62	11 (17.7%)	51 (82.3%)
Control Group n=68	8 (11.8%)	60 (88.2%)

NB: 50 Records on mentor database were incomplete.

Table 17: Constitution of groups by clinical grade

	GRADE G	GRADE F	GRADE E	GRADE D
Mentor Database, n=473	75 (15.8%)	16 (3.3%)	259 (54.7%)	123 (26.0%)
Experimental Group, n=62	9 (14.5%)	3 (4.8%)	34 (54.8%)	16 (25.0%)
Control Group, n=68	9 (13.2%)	2 (2.9%)	45 (66.0%)	12 (17.6%)

NB: 286 Records on mentor database did not contain clinical grade information.

Table18: Chi –square comparison of mentor database, experimental, and control groups by workplace, gender and clinical grade

COMPARISON	CHI-SQUARE	DEGREES OF FREEDOM	P
Workplace	12.91	8	>0.10
Gender	4.38	2	>0.10
Clinical Grade	3.72	6	>0.10

Summary of results of group comparisons

There were no statistically significant differences between the experimental and control groups, or between the study groups and the group of mentors within the mentor database.

10.2.2 Comparison of control and experimental groups

Demographic data from the RCT was compared

Table 19: Constitution of groups by years registered

YEARS SINCE FIRST REGISTERED					
	1-5	6-10	11-15	16-20	>20
Experimental Group (n=56)	17 (30.0%)	14 (25.0%)	13 (23.0%)	3 (5.3%)	9 (16.0%)
Control Group (n=64)	22 (34.0%)	20 (31.0%)	14 (21.0%)	4 (6.3%)	4 (6.3%)

NB: 10 subjects did not answer this question.

Table 20: Constitution of groups by experience of assessing students.

PREVIOUS INVOLVEMENT WITH STUDENTS			
	None	Some	Has been responsible for students
Experimental Group n=61	11 (18%)	26 (42.6%)	24 (39%)
Control Group n=67	12 (17.9%)	38 (56.7%)	17 (25%)

NB: two subjects did not answer this question.

Table 21: Constitution of groups by educational background

EDUCATIONAL BACKGROUND					
	ENB 998	SVQ Assessors Course	Professional Studies: Teaching, counselling, assessing	Teaching Qualification	None
Experimental Group n=62	0 (0%)	5 (8.0%)	9 (14.5%)	1 (1.6%)	47 (75.8%)
Control Group n=68	1 (1.4%)	3 (4.4%)	10 (14.7%)	3 (4.4%)	51 (75.0%)

Table 22: Constitution of Groups by professional qualification.

PROFESSIONAL QUALIFICATIONS				
	RGN (RN)	RGN plus other registration	RGN plus Diploma or Degree	RGN plus other registration, plus diploma or degree.
Experimental Group, n=62	25 (40.0%)	12 (19.0%)	17 (27.0%)	8 (13.0%)
Control Group n =66	30 (45.0%)	9 (14.0%)	22 (33.0%)	5 (7.5%)

NB: two subjects did not answer this question.

Table 23: Constitution of Groups by frequency of mentor contact with students in the workplace.

FREQUENCY OF CONTACT WITH STUDENTS				
	Constant	More often than not	Occasional	Never
Experimental Group n=58	22 (38.0%)	22 (38.0%)	14 (24.0%)	0 (0.0%)
Control Group n=68	24 (35.0%)	23 (34.0%)	18 (26.0%)	3 (4.0%)

NB: four subjects did not answer this question

Table24: Chi-square comparison: demographic data of experimental and control groups

COMPARISON BY...	CHI-SQUARE	D.O.F.	P
Years registered	3.18	4	>0.10
Assessment experience	3.14	2	>0.10
Relevant education	2.44	4	>0.10
Professional qualification	2.06	3	>0.10
Frequency of student contact	2.42	3	>0.10

Comparison of experimental and control groups by pre-test scores for anxiety, attitude and self-efficacy

Table25: Comparison of baseline scores for anxiety, attitude, and self-efficacy

MANN-WHITNEY U TEST	ANXIETY (PRE TEST)		ATTITUDE (PRETEST)		SELF-EFFICACY (PRETEST)	
	experimental	control	experimental	control	experimental	control
Mean Rank	70	61	66	65	62	69
U	1813		2047		1876	
P	0.17		0.78		0.28	

Tables 16 and 17 show no statistically significant difference between groups in any of the variables measured.

Comparison of baseline state anxiety of study groups compared with normative data.

Normative data published within the Manual for the State-Trait Anxiety Inventory [298] was used to compare the baseline state anxiety of study groups. This has been detailed in table 26.

Table 26: State Anxiety Means, Standard deviations, for working adults, students, and military recruits: Spielberger[298], pg14.

STATE ANXIETY	WORKING ADULTS		COLLEGE STUDENTS		HIGH SCHOOL STUDENTS		MILITARY RECRUITS	
	Male	Female	Male	Female	Male	Female	Male	Female
Mean	35.72	35.20	36.47	38.76	39.45	40.54	44.05	47.01
SD	10.40	10.61	10.02	11.95	9.74	12.86	12.18	14.42

Table 27: State anxiety means and standard deviations, for study groups.

STATE ANXIETY	STUDY GROUP (BASELINE)	STUDY GROUP (POST)
Mean	37.22	33.55
SD	11.61	8.51

Table 27 shows that the mean anxiety scores and standard deviation for the study group before the trial were similar to those of College Students, shown in table 26.

Summary of the results of group comparisons

There were no statistically significant differences between the experimental and control groups in any of the variables measured.

10.2.3 The analysis of anxiety, attitude, and self-efficacy data: RCT

Exploration of data.

Table 28: Experimental and control groups: mean, median and standard deviation for anxiety, attitude, and self-efficacy, pre and post educational intervention.

		PRE			POST		
		Mean	Median	S.D.	Mean	Median	S.D.
Anxiety	Experimental n=62	1.96	1.76	0.69	1.68	1.71	0.39
	Control n=68	1.77	1.71	0.48	1.67	1.54	0.46
Attitude	Experimental n=62	3.74	3.76	0.19	3.90	3.87	0.32
	Control n=68	3.74	3.72	0.21	3.78	3.77	0.19
Self-efficacy	Experimental n=62	3.34	3.28	0.56	3.79	3.79	0.49
	Control n=68	3.46	3.46	0.48	3.67	3.79	0.42

Anxiety diminished, as attitude and self-efficacy increased in both pre and post scores (table 28). In both attitude and self-efficacy, the experimental group had a better final score than the control group. Anxiety, much higher in the experimental group initially, converged with the level of anxiety in the control group post-intervention.

Table 29: Direction of movement in anxiety, attitude and self-efficacy scores during RCT

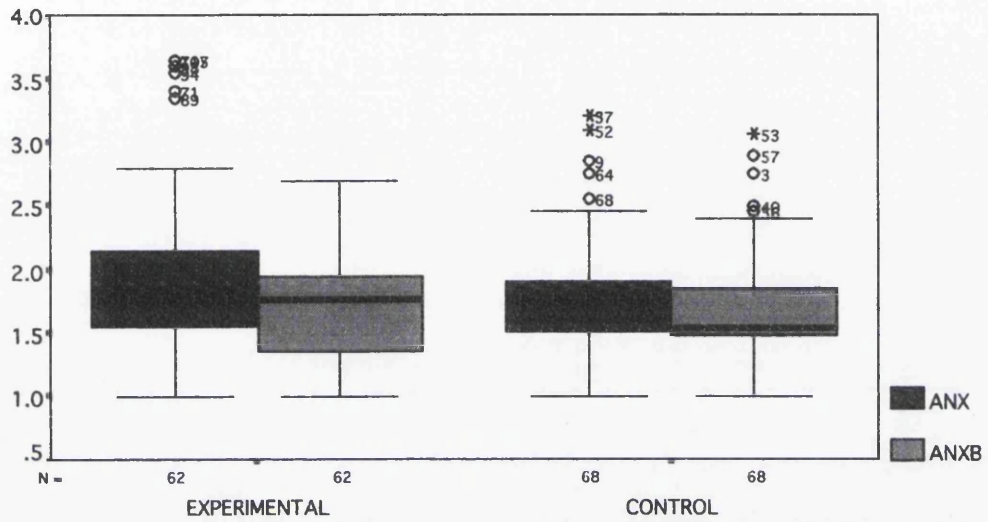
GROUP	SUBJECTS WHO...	ANXIETY	ATTITUDE	SELF-EFFICACY
Experimental n=62	deteriorated	19.3%	30.6%	21.0%
	improved	75.8%	61.3%	77.4%
	unchanged	4.8%	8.0%	1.6%
Control n=68	deteriorated	30.8 %	41.1 %	20.6%
	improved	55.9%	51.4%	76.4%
	unchanged	13.2%	7.3%	2.9%

Table 29 shows that the majority of subjects improved following the educational intervention in both the control and experimental groups. Some scores deteriorated, and some remained unchanged. Compared to the control group, the experimental group had a greater percentage of subjects showing improvement for all three variables.

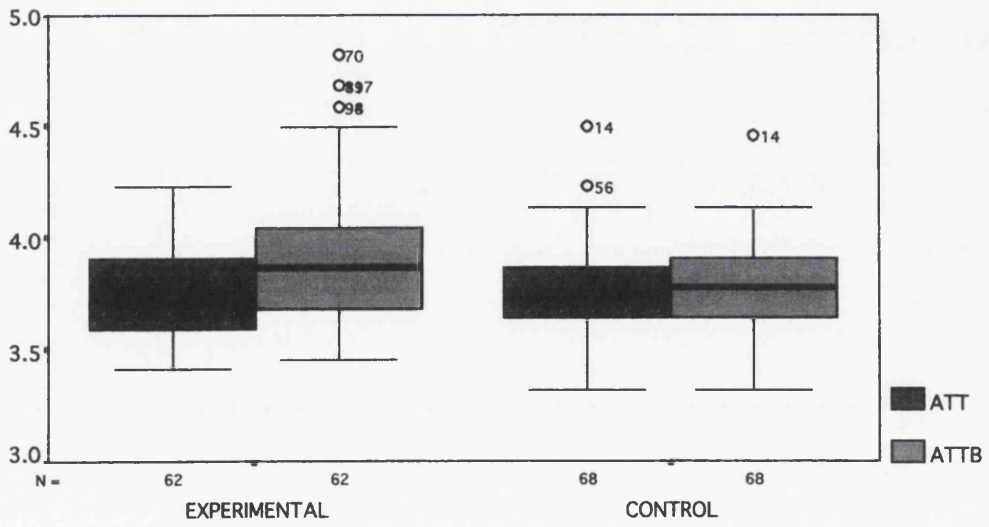
The following boxplot graphs show the distribution of the mean scores for anxiety, attitude, and self-efficacy pre and post educational intervention and reveal data to be

skewed, with a number of outliers. This influenced the decision to use the Sign test for statistical analysis.

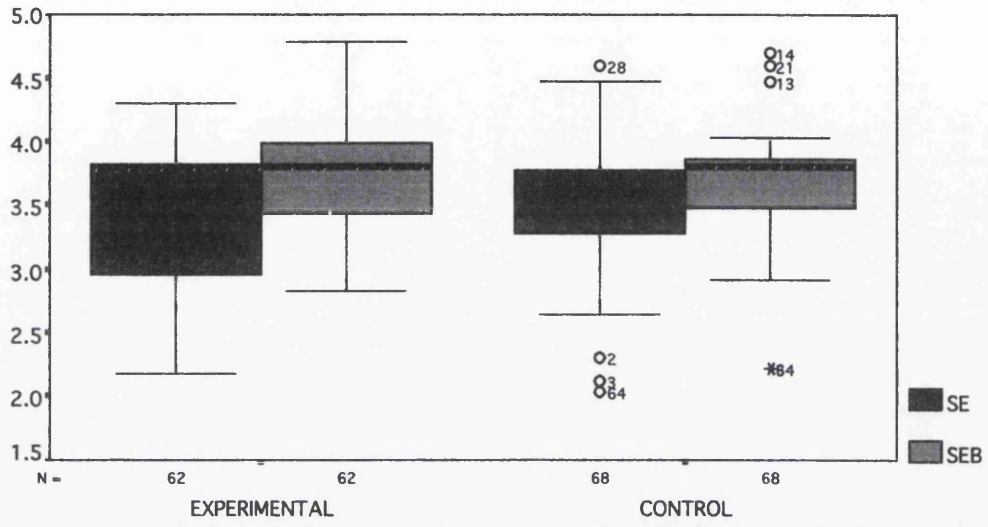
Boxplot (a) Anxiety: Anx (pre) and AnxB (post), for both experimental and control groups.



Boxplot (b) Attitude: Att (pre) and AttB (post), for both experimental and control groups.



Boxplot (c) Self-efficacy: Self-efficacy (pre) and SeB (post), for both experimental and control groups.



Analysis of data

Pre and post intervention scores for anxiety, attitude, and self-efficacy were compared for statistically significant differences using the Sign test. Comparison of final scores was analysed using the Mann-Whitney U test.

Table 30: Sign test: difference pre and post intervention for experimental and control groups.

	ANXIETY		ATTITUDE		SELF-EFFICACY	
	Z	P	Z	P	Z	P
Experimental n=62	4.43	<0.01	2.38	<0.01	4.35	<0.01
Control n=68	2.08	<0.05	0.75	0.44	4.55	<0.01

Table 31: Mann-Whitney U test: comparison of the final scores for anxiety, attitude and self-efficacy in the experimental and control groups.

	ANXIETY		ATTITUDE		SELF-EFFICACY	
	exp	control	exp	control	exp	control
Mean Rank	68	64	72	59	69	62
U	1980		1676		1883	
P	0.55		0.04		0.29	

In both groups, anxiety, attitude, and self-efficacy improved following educational intervention (table 30). Greater improvement was found in the experimental group where all variables showed improvement at the 1% probability level. This contrasts with the control group, where only self-efficacy improved at the 1% probability level, anxiety at 5% probability and attitude demonstrating little change. Mann-Whitney U test was used to compare experimental and control groups by final scores, as illustrated in Table 31, demonstrating no significant difference at the 1% probability in any of the variables. However, the experimental group showed better final attitude scores at the 5% probability level. This is important because, as was identified in table 28, the baseline attitude scores for both experimental and control groups were identical. This means that the null hypothesis in 7.2.3.2 relating to attitude can be rejected as a significantly greater improvement in attitude was found in the experimental group compared to the control group.

Finally, Kruskal-Wallis one-way ANOVA was used to compare changes in anxiety, attitude, and self-efficacy, between study groups.

Table 32: Differences between study groups pre and post intervention, in anxiety, attitude, and self-efficacy scores

	ANXIETY		ATTITUDE		SELF-EFFICACY	
	exp	control	exp	control	exp	control
Mean Rank	72	60	72	59	60	71
Chi-square	3.05		4.07		2.90	
P	0.08		0.04		0.09	

Confirming the results of the Mann-Whitney U test, the improvement in attitude was significant at the 5% probability level. Improvements in anxiety and self-efficacy scores were better in the experimental group; however, this only achieved statistical significance at the much weaker 10% probability level. These were insufficient grounds to reject the null hypotheses in 7.2.3.2 relating to anxiety and self-efficacy.

10.3 Follow-up interviews

Selecting "improved" subjects for interview

Subjects whose scores improved in all three sub-scales were identified using the programmable 'select cases' function of SPSS™. As can be seen from table 29, there was likely to be many whose scores come into this category. The formula, (anx-anxb>0.12) and (attb –att >0.12) and (seb-se>0.12), identified nine experimental and nine control subjects within these criteria. Three from both experimental and control groups were selected using the SPSS™ 'random' function. The remaining six cases were held in reserve in case of difficulty in contacting or interviewing those selected.

Selecting "deteriorated" subjects for interview

Subjects whose scores deteriorated in all three sub-scales over the course of the trial were identified. These were fewer in number (table 29) than those who showed across the board improvement, and the extent to which scores deteriorated was less. To select cases the formula (anx< anxb and att>attb and se>seb) was used. This produced three subjects in the experimental group, and six in the control group. Random selection was used to produce only three in each. Relaxing criteria for selection (dropping se>seb from the selection criteria) produced several additional candidates held in reserve in case of difficulty in contacting or interviewing subjects.

Interviews

These took place, between six and eight weeks after the trial, following the processing of data.

Interview protocol (appendix 22)

The questions related to the sub-sections of the RCT questionnaire: anxiety, attitude, and self-efficacy. Wording of questions applied to subjects from both the experimental and control groups. Data was in the form of quantitative, ordinal data from responses to questions, and qualitative data in the form of key points or issues arising from discussion.

10.3.1Follow-up interview results

Analysis of the group of twelve subjects selected for interview

A summary of responses to the RCT questionnaire has been included as table 33.

Table 33: RCT data, telephone interview group.

		CONTROL						EXPERIMENTAL					
Case number		62	48	30	40	53	61	78	127	86	76	81	84
Scores Pre & Post		Improved			Deteriorated			Improved			Deteriorated		
1. Gender		Female											
2. RN/RGN			√						√				
RN +2nd Reg								√					√
RN + Deg/Dip					√	√	√			√		√	
+ 2nd RN											√		
No response		√		√									
3. Experience		5	5	20	3	3	21	4	11	3	19	2	11
4. Grade		D	E	G	E	E	E	E	E	D	F	D	E
5. Involved in Assessment	None	√							√				
	Some		√	√	√	√	√	√				√	
	Total									√	√		√
6. ENB 998													
SVQ assessor				√			√						
PS I or II (assess.)											√		√
Teaching deg/dip													
None		√	√		√	√		√	√	√		√	
7. Workplace:													
Specialty													
Ward			√		√	√		√	√	√	√	√	√
Clinic or day hospital													
Nursing/Elderly hosp		√											
GP/ Health Centre				√			√						
8. Frequency of students													
Constantly				√		√				√	√		√
More often than not			√				√	√	√			√	
occasionally		√			√								
Change in RCT sub-scale scores													
Anxiety		-0.25	-0.40	-0.24	+0.70	+1.16	+0.14	-0.60	-0.55	-0.30	+0.15	+0.20	+1.00
Attiutude		+0.20	+0.27	+0.23	-0.14	-0.50	-0.05	+0.32	+0.27	+0.50	-0.05	-0.18	-0.27
Self-efficacy		+0.40	+0.39	+0.41	-0.13	-0.26	-0.10	+0.91	+0.83	+1.00	-0.09	-0.04	-0.06

Interview findings

In general, subjects whose scores improved during the RCT responded much more positively to the questions than those whose scores deteriorated, and the experimental group produced more favourable responses than the control group.

Quantitative data

The quantitative data from questionnaires was collated. These reflected the changes in score in each of the RCT sub-scales (table 34).

Table 34: Quantitative data from telephone interview subjects

Case Number	CONTROL						EXPERIMENTAL					
	62	48	30	40	53	61	78	127	86	76	81	84
Scores pre & post	'Improved'			'Deteriorated'			'Improved'			'Deteriorated'		
1. 'ability' to assess students												
Helped	√		√		√		√	√	√			
Unchanged		√				√				√	√	
Confused				√								√
Hindered												
2. Assessment – is it now more or less worrying?												
More				√								√
Less		√					√	√	√			
Unchanged	√		√		√	√				√	√	
3. Any change in 'approach' towards students & education?												
Improved	√							√	√		√	
Unchanged		√	√	√	√	√	√			√		√
Deteriorated												
4. Any Difference in 'Confidence' in ability to assess students?												
Improved	√		√		√		√	√	√			
Unchanged		√				√				√	√	
Deteriorated				√								√

Two subjects demonstrated strikingly similar patterns of response (Nos. 40 and 84), one from the experimental, the other from the control group. Yet, the reasons given were quite different. The subject from the control group (No. 40) asserted that the half study day was inadequate, and had not improved her state of preparedness to undertake assessment. However, subject No. 84 from the experimental group was disconcerted that her approach to assessment had been flawed and that she needed to improve her knowledge and skills: something she had yet to do. This phenomenon of quite different perspectives demonstrating the same pattern of data results is discussed more fully in section11.2.2 (p182), 'interpretation of anxiety results'.

Qualitative data

In addition to the specified questions, occasional probing and prompts were required to clarify responses. As demonstrated, this qualitative data has potential value in revealing much more than an analysis of quantitative data alone could provide. Subjects have been categorised according to control or experimental group, and whether responses improved or deterioration in the RCT. Responses have been summarised.

Category 1: 'improved' responses - experimental group

Subjects in this category responded very positively, and made spontaneous complementary comments about the program. Recollections from the half-study day were strong (as they were in the experimental group whose scores deteriorated). Two felt that it was a good introduction to assessing students, but one could not comment as she had yet to mentor her first student. One subject worried about using the computer initially, but quickly felt comfortable with it. One subject, who claimed that it enabled her to relate to what was happening and helped her feel closely involved in the case studies, found the film of actual students very useful. The organisation of the simulation was disconcerting for one subject at first, but she eventually became comfortable with it and focused on the content. She felt challenged to improve her score, and felt more confident when that happened. Two subjects were pleasantly surprised at the program. One saw it as an encouraging departure from traditional nurse education, the other thought that more

programs like this were needed to make nurse education more practice oriented and that this was a major step forward.

Category 2: 'improved' responses - control group

These subjects were positive on the benefits of the half study day. They enjoyed meeting with colleagues from different wards, and being given time to think about their mentoring role. However, details of the benefits of the half study day, in terms of learning how to assess students, were vague and poorly presented; there was a tendency to confuse the 'assessment' half-study day with others in the programme. This vagueness was shared with subjects in the control group whose scores deteriorated. One subject suggested that it was useful to have had the assessment document explained by the lecturer. These subjects generally did not seem concerned about assessing students; being taught by a university lecturer was re-assuring, as was 'knowing who to contact for future reference' (No. 62). One subject didn't feel terribly confident, explaining that 'it was cut and dried when it was explained in class' (No. 48) but she could not remember the details of what was taught. The attitudes of these subjects did not seem to be much affected. One questioned why her attitudes would change anyway (No. 30), one said her views were the same, another said that she was probably keener to get involved in teaching students than before. Similarly, 'confidence' in ability did not seem to be tremendously affected: one said that it would depend on the student, another that she felt a bit better about assessment, and more capable having had the documentation explained.

Category 3: 'deteriorated' responses - experimental group

These subjects did not enjoy the half study day. Two of the three did not like computers, and had little experience of them. One (No. 76) felt that she had been abandoned by the lecturer, who had not taken her lack of computing experience into account. Another (No. 84) could not concentrate on content, was worried about doing 'something wrong' and did not agree with the judgement of the computer in the simulation. Two individuals were concerned that they had 'missed out' by not being 'taught' in the traditional way (Nos. 76 and81). One was worried that her experience of assessing students contrasted markedly with the process shown on computer (No. 84). She felt that she had been 'doing things

wrongly' for several years. She was still coming to terms with that and had been reading around the topic since the session. The subjects generally felt that the half study day had little impact on their learning and on how they viewed student and nurse education. However, one admitted that the half study day taught her that computers seemed to be suited to some individuals more than others, and she felt that she could not be dismissive of the computer program simply because it had not suited her (No. 76).

Category 4: 'deteriorated' responses - control group

This group of subjects gave very similar responses. Their main complaints were in relation to educational jargon and volume of information and handouts, which they confessed, had not yet been read. Two subjects believed that they were no better prepared for assessing students than they were before the study day, which was of particular concern to one who was aware of her accountability (No. 53). On the positive side, one subject was less worried because answers to specific questions were provided, and the information was there if it needed to be referred to in future (No. 61). Attitudes to students were largely unaltered: there was no connection made between the half-study day and the subjects' approach to teaching and supporting students. Confidence in ability to assess students was little changed. One felt that she was still unprepared but no worse than before, one felt slightly better about it, the other a bit more concerned that she would still rely on 'gut feeling' rather than anything more structured and logical. She had hoped that this would have changed as a result of the half-study day.

Overall comments:

Individuals in the experimental group were very clear in their recollections of the half study day. By contrast, those in the control group found it difficult to give details of theirs. It may have been the novelty value of the program that distinguished the 'assessment' half-study day from others. The general impression from the interviews was that the main benefit of the traditional teaching session was the 'hidden curriculum': the social and collegiate nature of a gathering of professionals. The main benefit of the computer teaching session, however, was directly related to the impact made on subjects' perceived ability to assess students.

Chapter 11: Discussion and Conclusions

11.1 Summary

This chapter considers the results and findings of the study, their interpretation, and wider implications. As there were no systematic differences between groups, it was concluded that they were representative of the population of mentors at large, and that results could be generalised.

The 'novelty value' of the computer program may have positively influenced the effectiveness of the educational intervention for the experimental group. This was argued to be a motivational factor integral to multimedia, distinguishing it from traditional teaching methods. The failure of self–efficacy improvement to reach statistical significance between groups was considered a function of instrument design and small sample size. The reasons for raised anxiety in some individuals post – intervention appears to have been for different reasons in the control and experimental groups: those in the control group expressing frustration and disappointment whereas those in the experimental group had experienced a challenge to their long held beliefs in relation to their practise at assessing students.

It was concluded on the basis of the questionnaire results and telephone interview findings that the program improved mentors' assessment potential, most evident in the statistically significant improvement in attitudes. The research design was satisfactory in providing a representative sample to study and the 'assessment potential' questionnaire was user-friendly, producing data of value. In developing the computer program, the inexact science of performance assessment and the difficult nature of mentoring were affirmed. Problems with clinical assessment were difficult to completely resolve, but it was possible to achieve a balanced approach based on the available literature.

The computer package developed and tested in this study shows that assessment can be effectively taught to mentors using a 'stand-alone' package and that this approach should

be included in the range of methods used for teaching mentors how to assess. This is different to computer-mediated communication that typifies networked applications, where learning is a function of the vagaries of the student - teacher relationship. In stand-alone programs, an instructional design can be created with proven capability in achieving high levels of human-computer interaction that can deliver repeated quality assured learning in a way that computer-mediated communication approaches couldn't. There is potential for convergence, achieving the best of both worlds: computer-mediated communications courses have been introducing more 'e-tivities' in a departure from sole reliance on dialogue and stand-alone applications can link with readily updated web-pages and web-tools such as discussion groups. The second approach was taken with the latest version of 'A Mentor's Guide to Assessing Student Nurses'.

11.2 Discussion of results of the RCT and telephone interview follow-up

It was important to conclude how representative the study groups were, in order to decide the extent to which results could be generalised, and the validity of any conclusions that could be drawn. A variety of sources of data were used:

- Demographic data from the 'assessment potential questionnaire'
- Baseline measures of anxiety, attitude and self-efficacy
- Demographic data from a large mentorship database
- Normative data on anxiety by population sub-group

11.2.1 Systematic differences between study groups

As none of the analyses of data from study groups and mentor database yielded statistically significant results (tables 18,24 & 25), the null hypotheses stated in section 7.2.2 (p111) could not be rejected and it was concluded that there were no systematic differences between groups. This indicated that the study groups were homogenous and representative of the population of mentors at large. Interestingly, the mean baseline measure and standard deviation of state anxiety for the study group as a whole (table 27) were consistent with college students (table 26, appendix 28), as opposed to working adults. This is perhaps understandable as participants came to this study in the role of

students, although they were, in fact, working nurses on a half-study day. Not too much can be made of this, however, as Spielberger's normative data was collected prior to 1983 in North America [298]. It does, however, support the conclusion that the study group was a valid representation and not an aberrant sub-group of the population self-selected through a quirk in the sampling process. Although administratively cumbersome, it could be concluded that the sampling process was effective in minimising bias and producing comparable study groups from which results could be generalised.

11.2.2 The RCT and telephone interview follow-up results: discussion and conclusions

This section focuses on the analysis of results from each of the sub-scales of the 'assessment potential' instrument. These have been discussed and conclusions drawn. Finally, discussion is centred on the results of the RCT as a whole.

Changes in anxiety:

There were a number of outliers in both experimental and control groups, and the data looked skewed (boxplot a). This supported the decision to use non-parametric tests for statistical comparison, such as the Mann-Whitney and Sign tests. The baseline level of state anxiety in relation to assessment in the experimental group was noted to be high (table 28). However this was not statistically significant (table 25), and may have been coincidental. Both groups showed a significant reduction in mean anxiety post educational intervention (table 30), the experimental group showing significance at the 1% level of probability, the control only reaching 5%. The difference between groups was only significant at the 10% level of probability (table 32), and insufficient to reject the null hypothesis in section 7.2.3.2 (p111). in relation to anxiety.

The difference between groups, although not statistically significant, is interesting. It could be attributed either to confounding factors raising baseline levels of anxiety in the experimental group, or to the effects of the intervention. It is therefore worth considering the raised baseline measure of anxiety in the experimental group in more detail, as several factors could have contributed to this. In this regard, qualitative data from the follow-up telephone interviews has been useful.

Possible causes of raised baseline anxiety in the experimental group

The University setting for the experimental group may have contributed to this, as it has been shown that the SAI is sensitive to the conditions under which the test is administered [298]. Travelling to the city centre campus, parking, finding the location of the class and being part of an 'experiment' could easily have heightened general levels of anxiety in subjects. In addition, computer anxiety is a form of state anxiety that may have had an effect. This situation contrasts with that of the control group where the location was familiar, and the teaching conventional. In order to offset this, time was spent encouraging the experimental group to settle and feel relaxed before collecting baseline data. The time allowed for this, approximately twenty minutes may not have been long enough.

Crucial to the interpretation of results is the ability of the SAI to distinguish between general anxiety, and anxiety related to the assessment of students' clinical skills. Spielberger et al [298] explained that it is possible to generate anxiety in relation to specific hypothetical situations. Therefore, participants were asked to focus their thoughts and feelings on the assessment of students' practice, and all that this entailed, immediately before data was entered into questionnaires. This was performed for both the experimental and control groups, pre and post educational intervention: but there was no gauge as to how successful it had been. It could have varied from subject to subject and group to group. The researcher knew which of the groups was experimental and which was control, therefore sub-conscious bias may have been a contributory factor to how well this focusing was performed.

Post-intervention mean anxiety levels in experimental and control groups converged (tables28, and 31). It was possible that the statistically significant reduction in mean anxiety was caused by the teaching session coming to an end for both groups, rather than being related to the content of either session. As discussed in the literature review (5.5.4, p 68), reduction in anxiety can be counter-productive, if it signifies apathy. However, in the telephone interview subjects, there was no evidence of apathy in those whose anxiety

levels had decreased post-educational intervention. As the mean levels of anxiety following the trial was very similar for both experimental and control groups, whether this level indicated apathy was irrelevant.

Interpretation of anxiety results

In attempting to resolve the ambiguities in the interpretation of results, findings from the follow-up telephone interviews summarised in section 10.3.1 (p175) were useful. Subjects from the experimental group did not appear to have experienced anxiety in relation to the organisation, location, or conduct of the RCT. Spontaneous comments indicated an interest in finding out about an important aspect of their job that they knew to be difficult.

Those subjects whose anxiety levels had improved (worried less about assessment) attributed this in part to being better informed and more confident in their assessment role. One subject, No. 84 (tables 33 and 34), who was more concerned about student assessment following computer use, was a nurse with many years of experience, including previous experience of assessing students. She disagreed with the approach to assessment within the computer program and felt her judgement to be challenged by this. This may have been evidence of 'cognitive conflict' (Laurillard's concept of 'intrinsic feedback' 6.8.5, p101) which was not entirely resolved in the time between being involved in the trial and being interviewed. This is an interesting finding that was not encountered in the interviewed subjects of the control group.

Another interesting finding was that two subjects in the experimental group had increased anxiety immediately following the trial that improved to 'unchanged' at telephone interview. A similar pattern was evident in the control group. However, the reasons for the initial increase differed. One subject seemed frustrated that following the half-study day, she felt no better prepared. Unlike case no. 84, there was no evidence of 'cognitive conflict' contributing to the anxiety score. Therefore, although anxiety increased in some subjects of both study groups, this appears to have been a manifestation of frustration or dissatisfaction in the control group, whereas for the experimental group it appears to have

been a manifestation of cognitive conflict that can be beneficial to the process of learning (6.8.5, p101).

This difference in the effects of different teaching methods could be attributed to the depth of involvement that the participant had with the topic. The computer simulation required active participation, whereas classroom settings can induce passivity: the former beginning the process of learning, the latter often simply a process of information gathering to aid future learning that may or may not take place.

Conclusions from discussion on anxiety results

Several conclusions can be drawn from this discussion of the findings of the SAI.

Firstly, the experimental and control conditions were environmentally and organisationally different, potentially confounding the measurement of SAI. The ability of the researcher to focus subjects' thoughts on the hypothetical assessment situation before data collection could have varied from group to group. Allowing the group to be comfortable within the environment, and having only one researcher performing this function each time, probably prevented the validity of the data being compromised. This small study had fewer problems than large or multi-centre trials would experience, where environmental conditions and inter-researcher reliability can be major issues.

Secondly, responses of the telephone interview subjects have been crucial to interpreting data from the SAI. In view of this, it seems that the most appropriate interpretation is that:

(a) the raised baseline measure of anxiety in the experimental group was not statistically significant, and did not indicate anything other than a coincidental finding.

(b) the reduction in anxiety in the experimental group, although not reaching statistical significance at the 5% probability level, indicated a trend that would have reached statistical significance if the sample size had been larger.

(c) the production of 'cognitive conflict' a desirable educational outcome of learning, has a potentially confounding effect on the measurement of state anxiety. In this study, both state anxiety and cognitive conflict were related to the process of assessment, therefore the act of focusing on the assessment before completing the questionnaire was probably not capable of distinguishing one from the other. It is therefore possible that having information and skills about assessing students has significantly reduced anxiety, but that this has been moderated through inducing cognitive conflict.

Changes in attitude

The mean baseline measure of attitude was identical for the experimental and control groups as shown in table 28, although mean and standard deviation differed slightly, and several outliers were present as shown in boxplot b. The distribution of data appeared less skewed than the anxiety data (boxplot a), however, non-parametric tests were used because of the outliers. The attitudes of the control group improved slightly following the teaching session, while the improvement in the experimental group was much greater (table 28). As the baseline measures of attitude were identical, the statistically significant difference shown in tables 31 and 32 was sufficient to reject the null hypothesis 7.2.3.2, (p 111), in relation to attitude.

Possible causes of improved attitudes in the experimental group

This could have been due to the novelty value of the computer program, which many would have found stimulating. This novelty value has been recognised as a confounding factor in evaluative studies of computer use. Its presence has been supported by telephone interviews. Some subjects commented that the organisation of the afternoon and the use of the computer program came as a refreshing change. Most nurses have experienced 'behaviourist' cal programs in their education. The impression from the interviews was that the experimental group found that interactive, reality based programs represented a step forward for nurse education. This may be an effect similar to the 'positive, go-ahead image' that has been attributed to educators who use technology and may have affected the attitudes that subjects have in relation to nurse education generally. This contrasted

with interviewed control subjects who felt that their session was no different to the type of teaching with which they were familiar.

Subjects have been known to compare any new program with expectations based upon their norms and personal experiences. This has confounded evaluative research in CAL[374]. However, the interviews indicated that at least part of the improvement in attitude was due to subjects having a better understanding of the issues involved in student assessment, including the importance of their role in the observational assessment of performance. Subjects in the experimental group, whose attitudes were much more positive, commented that the program allowed them to fully appreciate the issues involved in assessment.

Conclusions based on the discussion of attitude results

The statistically significant improvement in the attitudes of the experimental group represented a positive benefit to arise from the use of the program. Differences between the control group and experimental group, such as the novelty value of using computers, have been identified. Whether these ought to be considered confounding factors or factors integral to the approach is debatable. The novelty effect of the multimedia program has clearly been one factor. The strong recollections that experimental group subjects had of their half study day, compared to the hazy recollections of the control group, was perhaps evidence of it. Not only was there novelty value in using a computer as opposed to classroom teaching, but there was novelty within the computer program itself, as each part of the program used different media and in different ways. It is this novelty effect, capable of sustaining motivation in learners and increasing time on task, which has the potential to influence learning. Therefore, far from being a confounding factor that requires a control, 'novelty' is integral to multimedia as an approach to learning that distinguishes it from the traditional. However, there were strong indications from the interviews that not all subjects held views based solely upon novelty value. In addition, computerphobia may have counterbalanced any positive attitudinal effects that novelty value may have had.

Changes in self-efficacy

The difference in mean baseline measure of self-efficacy for both experimental and control groups was not statistically significant. Data appeared skewed, especially in the control group, with a number of outliers noticed in boxplot c. It was therefore appropriate to use non-parametric tests of statistical significance. An increase in self-efficacy as a result of educational intervention was expected (5.6.6, p78). This was indeed the case, with post scores for both groups showing a statistically significant improvement, as shown in table 30. However, the difference between study groups only showed statistical significance at the 10% probability level (table 32), insufficient to reject the null hypothesis 7.2.3.2 in relation to self-efficacy.

Within the context of this study, the measure of self-efficacy has been free of confounding influences. The absence of a statistically significant difference between study groups is worthy of exploration.

Possible causes of 'no statistically significant difference' in self-efficacy between study groups

The data collection instrument may not have been sufficiently sensitive to adequately distinguish between groups. This feature of the instrument was not taken into account during its development. Similarly, because the nature of the items did not allow a mixture of negative and positive phrasing, this could have encouraged errors of central tendency which reduced the sensitivity of the instrument, and made it obvious to the user how to falsify 'good' responses. Perhaps the incorporation of a validity measure into the 'assessment potential questionnaire', such as a lie scale, as advocated by Spielberger et al [298], would allow data demonstrating a pattern of responses showing central tendency to be screened from the analysis.

Interpretation of self-efficacy results and conclusions

The telephone interviews generally supported the view that the experimental group had increased self-efficacy through subjects appreciating assessment issues by the effective

use of media and through simulated experience and feedback on performance. However, traditional educational methods can also make a difference to self-efficacy, through role modelling by the teacher, sharing of vicarious experiences through reflection and the verbal encouragement that occurs within a good group teaching session. Although this has perhaps not been as effective as multimedia in the teaching of the observational assessment of performance, it has registered a significant difference pre and post session.

The difference between the experimental and controls, although only significant at the 10% level, taken with the results of the telephone interviews gives a strong indication that self-efficacy can improve through using the program 'A Mentor's Guide to Assessing Student Nurses'. The failure to reach a statistically significant difference between groups has probably had more to do with instrument design, and sample size.

Methodological discussion

The initial study design was more suited to the situation that prevailed at the outset of the study, where mentorship preparation was undertaken in large groups within college accommodation. Disseminating mentor preparation to clinical areas made the logistics of performing a RCT much more complex and introduced several confounding variables, such as environment. However, this was the real world necessitating a pragmatic approach to the randomised-controlled trial. In addition to the confounding variables such as cognitive conflict and computerphobia, the results may have been influenced by differences between teachers in the control group, recruiting a volunteer sample, the social aspect of study days and other 'hidden curriculum' differences between the study groups. Individuals who applied through conventional means also joined those destined for the control group, and others who turned up on the day replaced some that cancelled. This was unavoidable and characterised this type of course [14], but made it essential that the study design did not involve complicated, protracted procedures requiring more than one attendance by subjects.

These were the complications and confounding variables known to the researcher. There will have been others that were not apparent but may have had an effect on the data collected, for example, the pre-test and post test design may have cued the subjects to the areas of evaluation, hence the areas of learning [406]. Theoretically, through selecting a large enough sample and through random allocation to experimental and control groups, these variables ought to have cancelled each other out.

The development of the measurement tool the 'assessment potential questionnaire' was time consuming and complex. It required to be based on indirect measures of the effectiveness of the program in achieving its intended outcomes. Of the three sub-scales developed, anxiety was more generalised and open to interpretation than the others. Attitude and self-efficacy were context specific and may have been sufficient by which to gauge the effectiveness of the program, especially given the confounding nature of cognitive conflict.

The usefulness of the questionnaire beyond this research has been untested, but there are indications that it may be applied more widely. For example, as a gauge of mentors' assessment ability it could be the basis for the feedback and support that has often been suggested [66]. It is more cumbersome than Darling's MMP, the basis of which is shown in appendix 6, or Cameron-Jones and O'Hara's model (table 5). However, it is more detailed, and could be computerised to give on-line feedback, perhaps through comparison with normative data.

The quantitative data collected with the 'assessment potential questionnaire' was insufficient on its own to fully understand the implications of using the computer program. This confirmed the value of the follow-up focussed telephone interviews. An additional benefit of delayed evaluation was finding no indication that subjects 'reverted' to their previous state of knowledge and skills prior to educational intervention. Some studies have shown observational assessors of performance were liable to quickly 'forget' what they learned [161, 218]. There was no sign of this in the telephone interview group. Responses largely reflected those taken on the day of the trial, post-intervention. The

small number of those interviewed, and the limited nature of the data collected could be criticised, although it was a pragmatic response to problems in tracking individuals who had not given full details of name and workplace. In addition, the interviews could have been more open instead of being focused. In-depth probing, tape-recording, transcription, and analysis as described in Chapple [463] may have provided more useful data, as much was lost by categorising comments and taking brief notes.

Sample size

The sample size for the study was an estimate based on the small pilot of the instrument before and after a traditional teaching session with prospective mentors, as described in section 9.1.2 (p147). This sample size was designed to detect statistically significant differences in all three sub-scales before and after, in which it was successful (table 30) but it was only sufficiently large to enable differences between study groups to reach significance at the 10% probability level in anxiety and self-efficacy. The indications are that if the sample size was larger, then differences between study groups would have reached statistical significance on all measures at the 5% level of probability. The data from this RCT could be used to accurately estimate the numbers required to reach statistical significance.

The value of anxiety, attitude and self-efficacy as outcome measures.

The anticipated effect on users' attitudes, anxiety, and self-efficacy, based on theoretical assumptions, was not forthcoming in all subjects. Student assessment is not simply intuitive, or necessarily comes through experience alone, but it is a skilled activity that can be tackled in a systematic way. Appreciating this, and being shown how to assess students with opportunities for practise ought to have reduced anxiety and had positive effects on attitude and self-efficacy. For some, this has been the case. For others, though, the experience was disconcerting. Initially confident individuals, forced to come to terms with the inadequacies of their long-established practice may well experience an increase in their anxiety, and adversely affect attitude and self-efficacy. This is more likely to occur in simulated, reality-based activities than in safer, traditional teaching environments. The value in using anxiety, attitude and self-efficacy as outcome measures

of the effectiveness of an educational process is questionable, unless the confounding measure, cognitive conflict, can be controlled.

Learning styles and choice of teaching method

The findings of this study could be interpreted as good, considering that within the experimental group, there were those who did not enjoy using computers and would not have chosen these as a method of learning, had a choice been made available on the day of the trial. The interviews suggest that these individuals performed poorly in the 'assessment potential questionnaire'. An important point to emerge from this is a number of individuals with similar views are likely to be present in any group. Estimates of computerphobia have been as high as 30% within the general population, as discussed in section 6.6.2 (p92). In this study, if the multimedia approach to learning was matched with users who were unafraid of technology and who had a congruent learning style, then, overall, it ought to have produced much better results. The program would have been more successful in achieving its educational objectives, than it would for a mixed group containing computerphobic individuals. This returns to the point that has often been made in evaluative studies of CAL. Some like it, some do not. The logical conclusion is that, for maximum effectiveness, it should be offered as a choice from several teaching methods, as discussed in section 6.7.1(p94).

The results imply that the educational approach and content of the program were reasonably effective. If more care had been taken with methodoligical issues, such as uncontrolled confounding factors, and sample size, then results better than 'no significant difference' may well have been achieved for all variables being measured. Nevertheless, in choosing whether the program should be used, or not, in a situation of 'no significant difference', other factors, such as convenience, and curricular integration ought to be considered. Cross-platform compatibility, cost, availability of computers and IT skills have been limiting factors in the use of computers in learning. However, it has been anticipated that most homes will have Internet access, and as a result of Dearing and Garrick reports [20, 21], students at universities will have easy access to computers, as too

will staff accessing the NHS-net. 'A Mentor's Guide to Assessing Student Nurses' has been designed to be a resource for many different curricula and contexts.

11.3 General discussion

11.3.1 The process of assessing student nurses' clinical practice

The literature review has acknowledged that competence is a multi-dimensional concept, and that no single measure is sufficient on its own [48]. Clinical performance has eluded attempts to demonstrate concurrent validity, suggesting that it is worthy of measurement in its own right. The contribution that the observational assessment of performance can make to competence assessment can be markedly improved if a systematic, rigorous approach is taken. In this regard, though, the nursing profession is no further forward in the search for Woolley's 'Holy Grail' [50]. The perfect process with an instrument that is valid and reliable is as elusive as ever. In the meantime, it is possible to have a sound process, a clear and precise instrument, and an assessor who is knowledgeable, confident and not overly anxious about the practicalities and consequences of passing judgement on another. This level of ability in an assessor can be achieved through a multi-dimensional approach to skills acquisition, as outlined, for example, in Goldstein and Sorcher's training model [220]. The principles are understood. These have underpinned the instructional design of 'A Mentor's Guide to Assessing Student Nurses'. It is recommended that the UKCC examines the preparation of mentors specifically in the assessment of practice, and acts to put standards in place. Several reports have called for courses to be lengthened and given accreditation, for example, Watson and Harris[13]: only then would mentors have the feeling that they had acquired a new set of competencies to equip them for their role.

Mentoring is undoubtedly a difficult task with conflicting roles. In some locations in the UK attempts have been made to avoid conflict by separating the assessment function from the mentor role. However, this may have unanticipated consequences, possibly producing conflict of a different sort. Unless the 'assessor' is working with the student closely, as a mentor ought to be, then it is likely that the assessment will focus on specific nursing activities in a reductionist way. This would reflect the acquisition of competence,

but not necessarily the development of a well-rounded performance. The mentor is already ideally placed to gather evidence relevant to assessment, weigh conflicting information, and come to a conclusion. The intrusion of another individual into this relationship would not necessarily be helpful.

Mentorship is a confusing title for a role, which in the UK does not conform to the ideal. Some have warned against becoming embroiled in terminology, but unless professional nurses have a common understanding about a role (such as mentor) or issue (such as professional competence), then confusion will proliferate. Perhaps one reason for the assessment of clinical practice being so problematic in post- Project 2000 has been due to mentors attempting to conform to a model of classical mentoring in which assessment plays little part. This would account for the inadequate preparation of mentors in the assessment of students. To avoid confusing people's expectations in relation to the role, perhaps it should be called something else. Pseudo-mentoring, although an accurate description of a role that encompasses assessment, coined by Morton-Cooper and Palmer [227], is a pedantic title that would be open to ridicule.

11.3.2 The program: construction and educational approach

The construction of the program was a complicated and time-consuming endeavour. In terms of technology, digital video was at the cutting edge in 1995, when production began. Hardware to cope with the requirements of memory and processor speed was expensive and scarce. More recently, such capability has been packaged cheaply for the home computer and video market, such as Apple's i-Movie ™. Recent improvements in capture and compression technology, had they been available earlier would have markedly improved the end product in terms of audio and video reproduction. Much more can now be achieved with less effort.

The truth in Mayes' [397] proposition that the person who learns most from a software package is the developer, not the student, became apparent through the course of this research. The underpinning theory of assessment required to be well understood and the concepts clearly identified, to enable these to be modelled and simulated. For example,

Benner's Domains of Nursing Practice had to be thoroughly understood in order to capture these on video, and to edit these. Fortunately, most video footage of nurses working in practice situations can be categorised according to these domains, demonstrating their appropriateness. It is doubtful if any individual could acquire the same depth of understanding of Benner's Domains from merely using the program.

The instructional design of this program, although straightforward and pragmatic, incorporated different learning theories. Cognitivist and behaviourist approaches have been used to present information in an organised way, and to give feedback. In the simulation, using a case study approach has utilised elements common to all problem based learning (PBL) models, by encouraging the user to explore, analyse and make decisions. Different media and a minimally directive approach have been integral to this approach. For example, people watching the same piece of video will perceive things differently. This has allowed elements of the program to be used and interpreted by individuals in their own unique, constructivist way.

The use of real staff, students, and patients has provided ecological validity that classroom settings and teachers alone cannot provide. The reality of this learning situation ought to encourage transfer of learning to practice. It is a fairly simple approach from a programming point of view, but one that is founded on sound educational theory. The relative success of the RCT is testament to this.

11.3.3 Evaluation

Much has been written about the merits and problems of different approaches to evaluative research. This research project, like many others, has demonstrated that both quantitative and qualitative approaches to research have benefits. The specific functionality of various parts of a program can perhaps be best evaluated through quantitative laboratory techniques, routine in the science of HCI. The temptation was to use this approach by employing Authorware™ monitoring and tracking facility to log keystrokes and links followed, in a laboratory setting with subjects. This objectively gathered data would have needed to be interpreted, however, losing its objective value.

The usefulness of this type of approach is very limited, as it gives no indication of the effectiveness of the program in relation to its outcomes, in terms of the acquisition of knowledge and skills in assessing students. Measures of effectiveness in relation to this 'outcome' were required. This applied both to formative and summative evaluation strategies. Some of the comments from lecturers and mentors during verbal protocol analysis were invaluable in providing an insight into what was effective and what was not.

Ideally, the best approach to gauging the effectiveness of the program would have been to compare how mentors assessed students, by direct observation, before and after using the program. As there are so many variables to contend with in clinical practice and the process so protracted, then this approach would have been unrealistic, justifying the development of an indirect measure. To an extent, the program was evaluated in an artificial situation: few mentors would use the program in a laboratory, and most would perhaps browse the www 'mentorweb' pages for up to date resources and information, in addition to using it to practise assessing students.

The strategy of evaluating 'A Mentors' Guide to Assessing Student Nurses' formatively, using mainly qualitative methods was successful in giving the project direction, and providing a version of the program that could be used in a RCT. An important point to make with formative evaluation is that it has to be started early in the life of a project, before 'ownership' develops, and criticism becomes harder to take. Evaluation could have been confined to formative evaluation, as it is in many commercial programs, where the marketplace is the final judge of a good program. However, in the absence of a marketplace in educational applications, the ability of a program to achieve its intended purpose with the intended target group is vital to determine, through summative evaluation.

The future

The future of computers in education has often been linked to the Internet, and the importance of computer-mediated communication. However, this research has shown that

there is real value in using carefully constructed and imaginative non-networked applications. This aspect of computer use in education appears to have diminished, and students may face the prospect of using a tool that allows access to information in an almost limitless way, with little expert educational guidance or structure. A balance between freedom and guidance needs to be achieved in multimedia programs, and indeed in web-based applications. This guidance can be complemented with appropriate educational approaches, which are varied in the structure that is imposed on the user's learning. Constructivist, cognitivist and behaviourist approaches can be used in harmony, whether it is on a web-based medium, or as a stand-alone package.

The prospect of education being de-schooled, through the extensive use of computer-mediated communication [474] consistent with Illich's vision has been thought a possible future scenario. There seems little prospect of this as long as students, as consumers, exercise choice in relation to educational method. A significant proportion of nursing students, for example, would not use computers for learning, preferring the social nature of face to face learning. It seems there is likely to be a market for campus based course for some time to come [23].

The rapidly changing educational backdrop, and organisation of mentor preparation during the production and evaluation of this program, made it absolutely essential that flexibility be accommodated within the structure of the program. This approach has been beneficial in the longer term. Following the recommendations in the Peach Report [39] and the publication of UKCC competencies[115], many institutions have re-written or updated their curricula, giving particular attention to the process of clinical learning and its assessment. For example, the Department of Nursing and Community Health, Glasgow Caledonian University, the biggest department of the largest Faculty of Health in the United Kingdom, has adopted an approach to assessment that utilises an adapted framework of Benner's Domains [90], with criteria developed from Bondy [196]and Steinaker and Bell [106]. This approach is common to all undergraduate degree and diploma courses. 'A Mentor's Guide to Assessing Student Nurses' has been re-formatted

for the PC (appendix 1c) and is being used as a resource for the preparation of mentors in the observational assessment of student nurses' performance.

During the period of this study, the Internet has grown from something that academics and governments found useful, into something that pervades the life of everyone involved in modern society. The CD-ROM was the only feasible mode of delivery for a sizeable program at the beginning of this project. Even now, streaming video on the Internet is in its infancy, and unreliable except in local, broadband networks. However, DVD is now a possible medium for future packaging, allowing much better video quality and length. As with the explosive availability of CD-ROM drives; the DVD has been primarily driven by entertainment applications such as full length, digital movies, available on television and computer. The shifting sands of nurse education reform and higher education globalisation provided an uncertain environment within which a program that was intended to use the latest technology and incorporate the implications of the most recent research was produced. A very difficult task for one individual who was simultaneously a full-time lecturer and spare-time student, to name but a few roles and responsibilities. Part-time PhDs in this field have become increasingly difficult.

Chapter 12: Recommendations

12.1 Summary

This chapter builds on the discussion and conclusions by identifying recommendations that ought to be of value to nurse managers, nurse educators and researchers.

12.2 Nurse managers

As it has been shown that a multimedia approach to educating nurses can be at least as effective as traditional means, then managers ought to make a cost-benefit analyses of different methods of in-service education to determine the best approach in different contexts. One advantage for managers in utilising multimedia would be consistency of educational delivery. The NHS Net and the clinically situated computers designed for nursing information and patient management systems are a resource that could be used for educational purposes. This would be cost-effective and would encourage computerphobic individuals to use computers more, with additional advantages for the individual and the organisation. Given the success of the liberal approach to education implicit within 'A Mentor's Guide to Assessing Student Nurses', it is recommended that a less traditional approach be taken to in-service education generally. This would give more control to employees as learners, and not necessarily focus narrowly on the acquisition of competence.

As it has been shown that the development of quality educational materials requires a team approach, it is recommended that nurse managers encourage the development of skills and knowledge in programming, multimedia, and instructional design in at least a small number of key clinical staff to enable nurses to lead the development of nursing software. Similarly, as it has been shown that multimedia programs are capable of representing complex activities for analysis, through video, audio and graphics, then it is recommended that more encouragement be given to nurses to become involved in this.

Capturing authentic clinical media to create a database for use in in-service and university education would be an exceptionally valuable learning resource.

12.3 Nurse educators

Given that the program 'A mentor's guide to assessing student nurses' has been shown to be at least as effective as the traditional teaching method, it is recommended that it be marketed, distributed, and made available to mentors and lecturers as a method of preparing mentors to assess students. This program could be packaged as an integral part of a level 3, 20 Scotcats module, or offered at level four, Masters or Honours level, with the 'case studies' as the focus for discussion. As the Internet continues to grow exponentially it could be anticipated that most nurses will soon have access at home, it is recommended that the program be converted to web pages, with the possibility of streaming video explored.

As the principles of clinical assessment have been well understood, it is recommended that the UKCC sponsor the development of a course that would meet mentors' needs. This would involve developing mentors' skills and knowledge, using varied methods, and simulated activities. In addition, a national assessment instrument tested for validity and reliability would have merit if it was commissioned by the UKCC. However, as context is vital to learning and assessment it could be argued that the UKCC has come as close as it reasonably can in outlining competencies that require to be assessed, while leaving it to individual institutions to create their own tools and processes *(11)*.

Given that mentors work with students and appreciate their strengths and weaknesses across the breadth of their performance, and that this is not possible by anyone other than the mentor, then it is recommended that the mentor and assessor roles be combined throughout the UK as they have been in Scotland.

Given that early CAL programs have been roundly condemned by educators as being capable of using nothing more than behaviourist methods, and that this program has

shown that a more holistic approach can be achieved through multimedia and careful instructional design, then it is recommended that nurse educators re-examine their attitudes towards using technology in teaching. As liberal philosophies of education are now proliferating to the extent that PBL is now a common approach within nursing curricula, then it is recommended that CAL programs be developed and used as student centred, fixed resource sessions, for which they are ideally suited. However, given that the rate of change in educational technology continues to outstrip the rate at which the population develops skills and knowledge to effectively utilise it, especially noticeable in nursing, then the use of computers for learning should only be provided as an option for students, as one method in a portfolio of approaches to learning.

As copyright restrictions prevent the use of pre-existing patient focussed media, such as that developed for use in the BBC programme 'Casualty', then it is recommended that partnerships between media publishing companies and educational institutions be explored.

12.4 Researchers

Given that the concept of 'cognitive conflict' was found to confound anxiety measurement using the SAI, it is recommended that its use within this context ought to be with caution, and the relationship between these concepts needs to be explored in further research. Based on the findings of the follow-up interviews, it may well be that educational intervention has effects on an individual's anxiety that differs dependent upon their experience and knowledge of the subject. This would predict that educational intervention for inexperienced individuals may reduce anxiety in relation to a subject, whereas experienced individuals, mistakenly confident in their own abilities, suffer from increased anxiety as a result of educational intervention.

As there has been potential variability in the process of focusing groups on hypothetical situations prior to the collection of data with the SAI, for larger or prolonged studies, it is

recommended that this be a much more structured process, involving perhaps video recorded instructions that would be possible to show repeatedly.

Given that the 'assessment potential' questionnaire has been fairly successful as an indicator of a mentor's predisposition to effectively assess students, and that attitude and self-efficacy have been less likely to be confounded than anxiety, it is recommend that a refined tool be developed focusing specifically on these attributes. The possibility of adjusting the sensitivity and specificity of the self-efficacy instrument ought to be explored. As self-efficacy measuring instruments also allow users to easily falsify 'good' responses, and facilitate errors of central tendency, it is recommended that a lie scale be incorporated to distinguish conscientious data entry from casual data entry.

Given the protracted time-scale for development of quality, formatively evaluated learning programs and the rapidity of change in nursing and health care, it is recommended that projects with a limited life expectancy ought to be sufficiently resourced from the outset to reduce development time in order to get as much return from the investment as possible.

As the program 'A Mentor's Guide to Assessing Student Nurses' successfully used the technique of comparing user input with 'expert judgement' as an alternative to interpreting free text entry, then it is recommended that this approach be used in other applications. As higher levels of interaction can best be achieved by allowing users to enter text freely, it is recommended that a way be found to use a nursing thesaurus and language classification system within educational applications to allow the program to interpret and respond appropriately.

Appendices

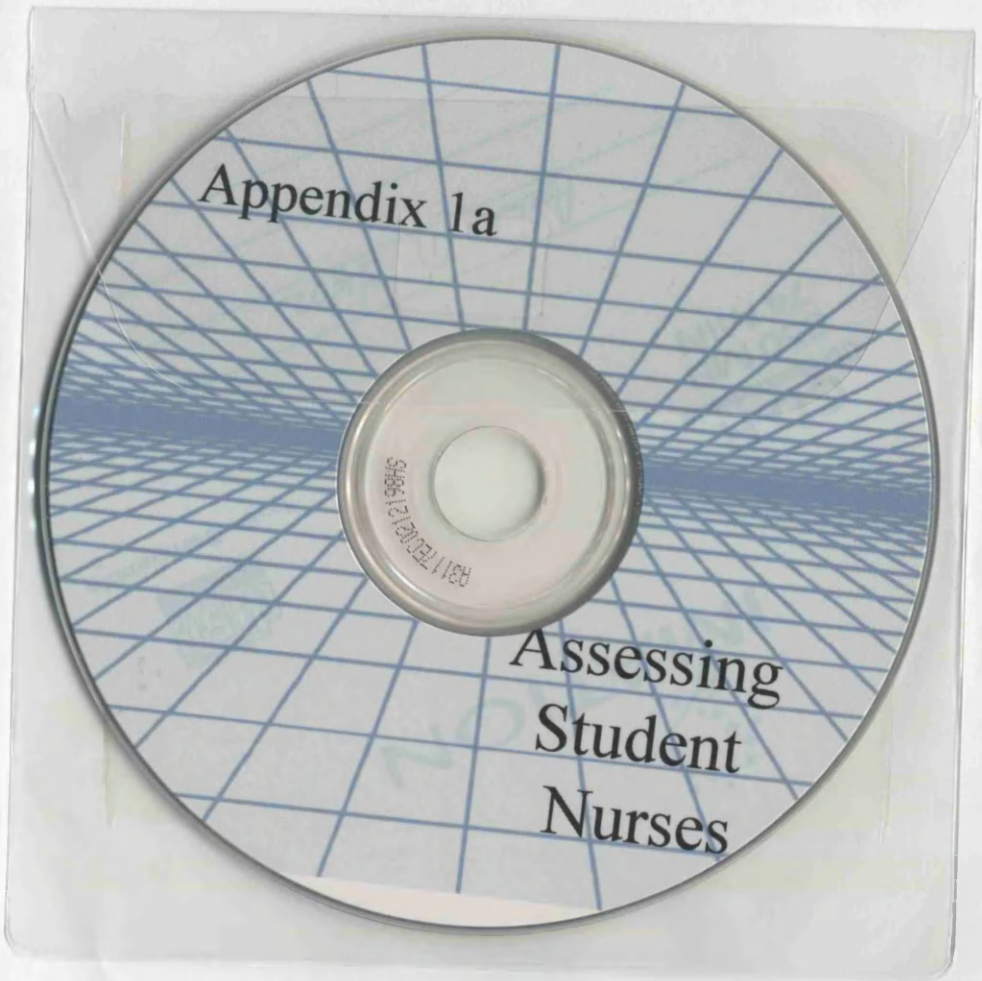
Appendix 1: Development versions of "A Mentor's Guide to Assessing Student Nurses"

Appendix 1a:Prototype used in formative evaluation

"Assessing Student Nurses"

This version is Mac only.

Please read and follow the instructions contained within the CD-ROM 'Read Me' file. Once files have been installed in the appropriate place on the hard disc, the program should run satisfactorily from the CDROM although better performance could be gained through copying the CDROM to your hard disc, and running it from there.

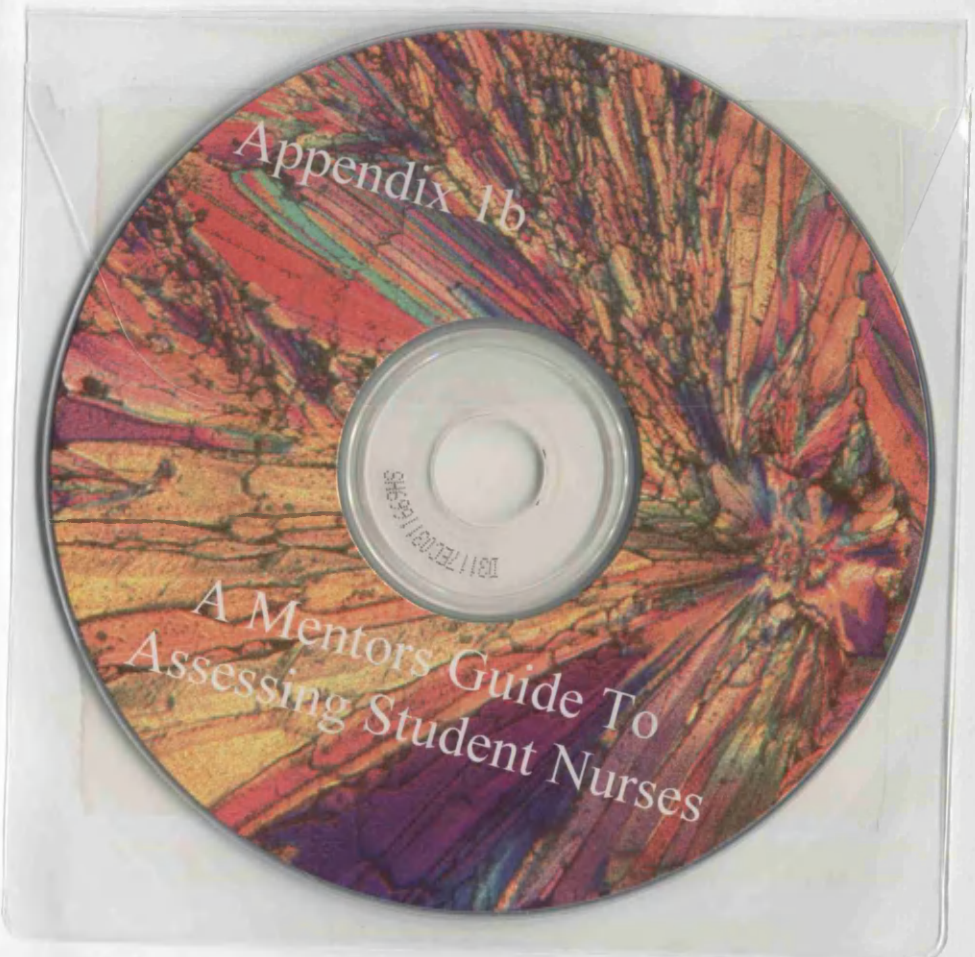


Appendix 1b: Randomised Controlled Trial version

"A Mentors Guide to Assessing Student Nurses"

This version is Mac only.

Please read and follow the instructions contained within the CD-ROM 'Read Me' file. Once files have been installed in the appropriate place on the hard disc, the program should run satisfactorily from the CDROM.



Appendix 1c: PC version

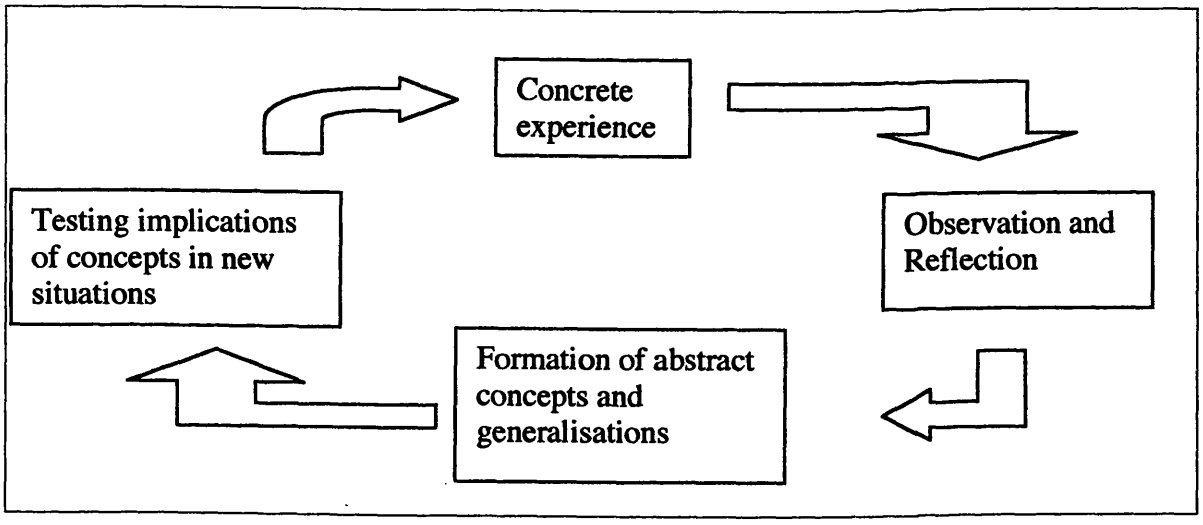
"A Mentors Guide to Assessing Student Nurses"

This version is PC only.

It has been developed to allow wider access to the program, and to relate to the 'mentorweb' pages more closely. Some persistent bugs of earlier versions, such as not being able to close an audio file once open, have been solved.

Open the CDROM in your computer, then click on the 'Mentor' or 'Mentor exe' icon, ignoring the others. This version runs entirely from disc.

Appendix 2: Kolb's experiential learning model [102]p.21



Appendix 3: Guilford's 6 sources of rating error *(166)*

- ERROR OF LENIENCY
- ERROR OF CENTRAL TENDENCY
- HALO EFFECTS
- LOGICAL ERROR
- PROXIMITY ERROR
- CONTRAST ERROR

Appendix 4: Bondy's 5 point criterion referenced scale *(196)*

CRITERIA FOR CLINICAL EVALUATION			
Scale Label	Standard procedure	Quality of performance	Assistance
Independent	Safe Accurate Affect And Effect each time	Proficient, co-ordinated, confident. Occasional expenditure of excess energy. Within an expedient time period	Without supporting cues.
Supervised	Safe Accurate Affect And Effect each time	Efficient, co-ordinated, confident. Some expenditure of excess energy. Within a reasonable time period.	Occasional supportive cues
Assisted	Safe Accurate Affect And Effect each time, most of the time	Skilful in parts of behaviour. Inefficient and uncoordinated. Expends excess energy; within a delayed time period.	Frequent verbal and occasional physical directive cues in addition to supportive ones.
Marginal	Safe, but not alone. Performs at risk. Accurate – not always. Effect & affect: occasionally	Unskilled; inefficient. Considerable expenditure of excess energy. Prolonged time period.	Continuous verbal and frequent physical cues.
Dependent	Unsafe Unable to demonstrate behaviour	Unable to demonstrate procedure/behaviour. Lacks confidence, co-ordination, and efficiency.	Continuous verbal and physical cues.

Appendix 5: The principles of assessing nursing practice: Wilkinson [166]

- Developing communication and liaison with link teachers
- Having a knowledge of the students' learning objectives
- Allowing each student a settling in period
- Avoiding stereotypes by focusing on the individual student
- Using documentation in the way that it has been designed
- Seeking opinions of others in the ward team
- Identifying typical performance, rather than exceptionally weak or strong examples from the student's repertoire
- Acting upon unsafe practice immediately
- Assessing cognitive, affective and psychomotor aspects of performance
- Acknowledge the context of the assessment
- Providing frequent feedback to the student
- Supporting judgements of the student's strengths and limitations with evidence
- Promoting reflection through listening
- Challenging the student, accompanied by support throughout the assessment process
- Reviewing performance regularly with reference to agreed criteria, and documenting feedback.

Appendix 6: Darling's 14 mentoring activities [225]

SUPPORTER	FEEDBACK-GIVER
COACH	EYE-OPENER
MODEL	DOOR OPENER
ENVISIONER	IDEA BOUNCER
ENERGISER	PROBLEM SOLVER
INVESTOR	CAREER COUNSELLOR
STANDARD-PRODDER	CHALLENGER

Appendix 7: Typology of computer interaction: Somekh (382)

- **Type A interactions**: recognition. Text-dependent, require matching of superficial features of information; non-productive
- **Type B interactions**: recall Text-dependent; superficial engagement of student with content; reproductive; manipulation of syntax and logical features of text.
- **Type C interactions**: reconstructive understanding or comprehension: text-independent but discourse dependent; involve schematic interaction with content; reconstructive, productive, involving the comprehension of statements, concepts or principles.
- **Type D interactions**: global reconstructive or 'intuitive understanding: experiential learning; discourse independent, focus on structure of discipline; student as master of discipline, problem-solving.
- **Type E interactions**: constructive understanding. Student creates fields of knowledge; discipline independent; exploratory; problem finding; domain-dependent.

Appendix 8: List of those involved in filming and media production.

FILMING INVOLVED:

18 PATIENTS

7 STUDENTS

12 STAFF NURSES

4 AMATEUR ACTORS

1 CAMERAMAN

1 SOUND TECHNICIAN

AUDIO CLIPS WERE NECESSARY, USING

12 SCRIPT READERS

1 AMATEUR ACTOR

Appendix 9: Letter from Churchill Livingstone

Direct Line : +44 (0) 131 535 1760
Direct Fax : +44 (0) 131 558 3171
E-Mail : peters@edinburgh.rsh.pearson-pro.com

Churchill Livingstone
Robert Stevenson House
1-3 Baxter's Place
Leith Walk
Edinburgh EH1 3AF
United Kingdom

Telephone: +44 (0) 131 556 2424
Fax: +44 (0) 131 558 1278

Direct Line

D

Charles Dougherty
Glasgow Caledonian University
Stobhill Campus
300 Balgrayhill Road
Glasgow G21 3UR

5th December 1996 Sent by Fax : 0141 558 9069

Dear Charles

First I must apologise for the disappointingly long time it has taken us to respond after our promising first meeting. If it has been disappointing for us, I am sure it has been doubly so for you and cannot have enhanced your view of publishers' efficiency. I can only assure you that this has been an exceptional delay, even taking into account the summer recess.

Our current view I am afraid is that we feel we cannot proceed with the project. I say "current", because it is possible that this view may change after another passage of time. We have experienced - like most publishers - mixed results with our first wave of multimedia publishing. We are about to embark on further, structured research into the market for educational multimedia in the UK and beyond. It is clear that several departments, like your own, have initiated multimedia projects for the nursing, medical and healthcare markets. What is considerably less clear is the investment required to complete these projects and bring them to market, and - perhaps even more important for publishers - precisely what that market would be. For the foreseeable future, the individual student market appears very price-restrained, whilst the institutional market has complexities in the area of curriculum "match", budget holding and decion-making patterns. The foreseeable future however does not stretch very far!

Our aim over the next 3-6 months is to form a clearer picture of the type of materials a) under development around the UK, the rest of Europe and to an extent North America; and b) which would be used, and by whom, if they were available. During the course of this research we hope to talk in more detail with multimedia authors, designers and producers in the university sector, and it is entirely possible that we will wish to re-visit projects which we for the present have to put on hold.

To conclude, I would dearly like to be in the position - especially after such time - of saying that we can now move forward. What can be said is that this is a reflection of our view of the state of the market, rather than of the project itself, and that we hope to remain in touch as our research progresses.

Yours sincerely

PP

Peter F Shepherd
Director
Nursing and Allied Health/Medical Education

Appendix 10: Consent form - patient

Filming nurses at work : Patient Consent Form.

Name .....

Address .....

.....

Hospital .....

Ward .....

Date .....

I hereby give permission for specific aspects of my nursing care to be filmed on video. What this involves has been clearly explained to me, and my consent is provisional on the understanding that the video film will be used for educational purposes only.

Signature.....

Please print name and address overleaf, if different to that of the patient. Please state your relationship to the patient.

Appendix 11: Consent form - student

Consent for Permission to Acquire Video

Dear Student Nurse..... Date.....

Thank you for volunteering to have aspects of your practice videotaped. It may be possible for us to use some aspects of this for future educational use. Naturally, we require your permission for this.

The exact portions of the video will be shown to you, and your consent sought at that time. At any time afterwards, and for whatever reason, you have the right to request that we do not use the video, which we will respect.

Please confirm that you understand this request, and either give, or withhold your consent.

I *give/do not give consent to this request.

Signed.....

*Delete as necessary

Thank you,

Charles Docherty
Nurse teacher

Appendix 12: Patient information

Patient Information - Filming Student Nurses at Work

There will shortly be members of staff from the College of Nursing in this ward, taking videos of student nurses.

They may want to film some aspects of your care - if this is so, you will be asked for your consent. Every effort will be made to ensure that your privacy is respected.

If you are at all worried by the thought of being recorded on film, then you have every right to refuse. The college staff will respect this without further discussion.

Should you give consent and then wish the film not to be used, then it will be destroyed.

If you do give consent and the film is later judged to be an invasion of your privacy then it will be destroyed.

A typical piece of filming would be of a nurse taking your blood pressure, or giving you your tablets.

Be assured that any video film will be used purely for educational and professional purposes, to the benefit of future nurses and patients.

Your co-operation in this will be greatly appreciated.

Charles Docherty,
Nurse Teacher
Glasgow College of Nursing and Midwifery

Appendix 13: Victoria Infirmary NHS Trust ethics approval letter

THE VICTORIA INFIRMARY
NHS TRUST

Our Ref: KC/EF.

Phoning ask for Direct Line - 0141 201 5411 - Dr Cochran's Secretary

Dictated 17.11.94
Typed 17.11.94

Mr C Docherty
Nurse Teacher
Greater Glasgow Health Board
South Division
2001 Govan Road
GLASGOW G51 4TF

c.c. Mrs M Hilton
Secretary
Ethics Committee
Victoria Infirmary NHS Trust

Dear Mr Docherty

TEACHING MENTORS - HOW TO ASSESS STUDENT SKILLS - A MULTIMEDIA APPROACH

It was nice to meet with you in the presence of my colleagues Dr Cowan and Ms M Campbell Smith. You detailed to us the reasoning behind your study, the clinical skills recorded to be used in the furtherance of the maintenance of nursing standards. At the moment the clinical skills involved would not be those that would cause embarrassment to a patient or relatives and we would ask that if such techniques were to be recorded that you inform the committee.

There were concerns regarding the distribution of the material and the ability of it to be kept "private" and for teaching purposes. You explained that it would be videoed, put onto hard disk and only from there would it be published.

We think it appropriate that the patient is informed of the nature of the teaching material and it would be worthwhile giving to relatives a copy of the information sheet. We suggested to you that the Clinical Directors of the 2 major Directorates involved, ie Dr S D Slater and Mr J K Drury be informed of your intention to video patients under their particular Directorates care. It would be a courtesy if included in this letter was a request that clinical colleagues be informed and were happy that their patients be involved.

We would be keen to have some feedback from you on the success of this interesting venture. I would think it appropriate that you go ahead and I hope you can meet the timescale involved. Under separate cover you will receive further information for your records from Mrs Hilton.

Yours sincerely

KENNETH M COCHRAN
CONSULTANT PHYSICIAN
& GASTROENTEROLOGIST

Victoria Infirmary,
Langside Road, Glasgow G42 9TY.
Telephone: 0141-649 4545. Fax: 0141-649 2206

Appendix 14: Data-entry and processing for "Simulation"

The input to the computer at this stage is a value for each of the domains, d1-d6.

For example, the data set (d1-d6) for a user's assessment can be shown in diagram 1.

Diagram 1: Data values for variables d1-d6, following data input "d" to grid

		LEVEL OF SKILL				
	DOMAIN	1	2	3	4	5
d1=3	HELPING			d		
d2=2	TEACHING		d			
d3=5	THER. REGIMES					d
d4=3	DIAGNOSTIC			d		
d5=4	ORGANISATIONAL				d	
d6=4	QUALITY				d	

Data Processing

In order that meaningful feedback can be given, the computer has a record of the correct level of skill in each domain, arrived at by a panel of 'expert' mentors who previously analysed the information available to the user. This is shown in diagram 2.

Diagram 2: Data values for variables e1-e6, as arrived at by a panel of experts, and stored within the computer program.

		LEVEL OF SKILL				
	DOMAIN	1	2	3	4	5
e1=2	HELPING		e			
e2=2	TEACHING		e			
e3=4	THER. REGIMES				e	
e4=3	DIAGNOSTIC			e		
e5=4	ORGANISATIONAL				e	
e6=5	QUALITY					e

The computer compares the user's data input (d) with the 'correct' version (e). The number of times the user and experts agree can vary between 0-6, and gives an indication

of the accuracy, or precision of the user's judgement. Similarly, the number of times the user is lenient (d>e), or harsh (d<e), can vary from 0-6. The relative values of these variables indicate bias and precision. The range, from maximum leniency to maximum harshness can be as great as 8: the higher this value, the higher the bias, and the lower the precision. The number of times the user agrees, is lenient, or harsh, and the range within which the user's judgement varies can thus form the basis of feedback. This information has been condensed into a 4 digit nominal number, the **alhr** (accurate, lenient, harsh, range) number that can be stored and used by the computer.

Producing the alhr number

Step 1: Identifying the variation between user and expert scores.

For each domain, the variation (v) is calculated by subtracting the expert value, the 'e' variable, from the user's data, the 'd' variable. In the example given,

V1 = 3-2=1

V2 = 2-2=0

V3 = 5-4=1

V4 = 3-3=0

V5 = 4-4=0

V6 = 4-5=-1

If V=0, then user and expert agree on the level of skill for that domain.

If V>0, then the user has been lenient

If V< 0, then the user has been harsh

The difference between the maximum and minimum values for harshness and leniency is the range.

Step 2: quantifying "a", the agree factor

For V1-V6, if V=0, then a=a+1 : next V

Multiply a by 1000. In this example, a=3x1000 = 3,000

(The reason for this multiplication will become clear in step 6)

Step 3: quantifying the leniency factor, "l"

For V1-V6, if V>0 then l=l+1, next V.

Multiply l by 100. In this example, l=2 x100=200.

Step 4: quantifying the harshness factor, "h"

For V1-V6, if V<0, then h=h+1, next V.

Multiply h by 10. In this example, h=1x10=10.

Step 5: quantifying the range, "r"

Identify maximum value of V, *Vmax*

Identify minimum value of V, *Vmin*

r= *Vmax-Vmin*

If r>5, then r=5.

(It is possible for the range to be as high as 8, but in terms of what one concludes from this, and the advice one gives, it is no worse than having a range of 5.)

Step 6: constructing the alhr number.

This is an ordinal number, obtained by adding the values for "a", "l", "h", and "r".

In the example given, the alhr = 3,000 + 200+10 + 2 = 3212.

(If no multiplication had occurred in steps 2,3 and 4, then the alhr could not have been constructed to represent each component: adding 3+2+1+2=8!)

Given the constraints of the data entry grid, and limiting the 'range' to a maximum value of 5, there are 109 possible alhr combinations. These situations have been ranked in order of 'least corresponding to expert judgement' to 'most corresponding to expert judgement', on an ordinal scale of 1-100. In some cases, the distinction between the types of situations that different alhr numbers represent was too fine to make any meaningful difference: in these cases, alhr numbers were ranked together. In addition to ranking, a feedback statement was prepared, addressing the type of situation that each alhr number (or group of alhr numbers) represented (in terms of being accurate, lenient, etc). In good scores, praise was given; in moderate scores, more practice was suggested; and in poorer

scores, advice to study the "concept demonstration" part of the program was given. Strongly advising the user that they were not yet ready to practice was reserved for the poorest scores, with the worst possible score receiving the comment "you didn't really try, did you?". These statements were constructed in collaboration with a senior lecturer in adult nursing, who was familiarised with the objective of the exercise.

In the example above, the alhr number of 3212 receives the following feedback:
On a scale from 1 to 100, you score 68.
"You are correct in your assessment of three domains of nursing practice, misjudging only marginally in the remainder. There is inconsistency in your judgement of the student, however, underestimating performance almost as often as overestimating. Your reliability and accuracy should improve with more practice. You are on the right lines: further practice should result in a much-improved score."

By comparison, an alhr number of 4112 (that is, 4 entries in agreement with the expert panel; one lenient, and one harsh judgement, within a minimal range of 2), produces the feedback statement:
On a scale from 1 to 100, you score 82.

> *You are correct in your assessment of four domains of nursing practice, misjudging only marginally in the remainder. There is inconsistency in your assessment, however, both underestimating and overestimating student performance, even if only marginally. This should improve with more practice. This is a very good attempt.*

On the next screen, an indication of how well the user has progressed is shown on an animated scale, the pointers indicating current score and previous score alongside each other. The ALHR number is then stored in the user's file on the computer's hard disc for future comparison.

Appendix 15: Schedule of questions for formative evaluation

These questions were used as prompts during the one-to-one sessions with experts. Not every question was used, as this depended upon where the expert chose to go in the program.

Log - in

Are the instructions clear?
Is there any room for misunderstanding?

Introduction to the program

Does the text tell you what you think you need to know at this stage in the program?
Is there anything unclear about the information?
Is it obvious to you what the program is intended to achieve?
Is it obvious to you how the program is structured?
Is the information accurate?
Is there any information superfluous for an introduction?
Is the content complete?

Navigation instructions

Are these sufficient?
Are these clear?
Is there any ambiguity?

Model (of the assessment process): general

Does the introductory text tell you what you think you need to know?
Is there anything unclear about the interface?
Is it obvious to you what is to be achieved in this part of the program?
Do you think it will be obvious to mentors what is to be achieved here?
Is the model an accurate reflection of current practice?
Are the steps as described in logical sequence?
Is the model diagram an effective "advanced organiser" for the information contained therein?
If you could change just *one* thing, what would it be?
Is the model a genuine representation of reality?

Model: constituent parts
Hot spot (link) key word.....
Sub-section:
Specific question asked (if any)

General questions to ask:

Is the information clear, accurate, and current?
Would mentors want to know about this?
Is the information superfluous?
Is the content complete?
Is there sufficient variety of icons/page layouts?
Are the icons consistent with the content they represent?
Does this section appeal , visually?

Concept Demonstration

Does the introductory text tell you what you think you need to know?
Is there anything unclear about the interface?
Is it obvious to you what is to be achieved in this part of the program?
Do you think it will be obvious to mentors what is to be achieved here?
Are the icons consistent with the content they represent?
Do you agree that these chosen concepts are the most important for student assessment?
Would mentors need to know about these?

Domains of Nursing (Sampling Performance)

Does the introductory text tell you what you think you need to know?
Is it obvious to you what is to be achieved in this part of the program?
Is the information clear, accurate, and current?
Would mentors want to know about this?
Is the information superfluous?
Is the content complete?
Is the navigation suited to this part of the program?
Do you think it will be obvious to mentors what is to be achieved here?
Is there anything unclear about the interface, at any point?
Are the photographs representative of the respective domains?
If you could change just *one* thing, what would it be?
Is there sufficient variety of presentation of information?
Does this section appeal to you, visually?

Domains

self-assessment

Does the introductory text tell you what you think you need to know?
Is it obvious to you what is to be achieved in this part of the program?
Does this section achieve its aims?
Video Quality
Is the video clip essential?
Could the aims of this part of the program be achieved without the video clip?
Is the feedback adequate?
Is it essential that further details of errors be given in feedback?
Do you agree with the "correct" response?
Is there a sufficient number of video clips?
Does this section appeal to you?

Levels of Skill (Judging Performance)

Introductory video:
Is the video clip essential?
Does it adequately set the scene for this part of the program?

Information text screen
Does this text tell you what you think you need to know?
Is it obvious to you what is to be achieved in this part of the program?
Is the information clear, accurate, and current?
Would mentors want to know this?
Is the information superfluous?
Is the content complete?
Are the 'hotspots' (links) Apparent?
(should there be more of these?)

Exploration of different levels of skill

Do you think it will be obvious to mentors what is to be achieved here?
Are the additional navigation instructions for this part of the program superfluous/helpful?
Is there anything unclear about the interface, at any point?
Are the icons representative of the respective areas of content?

For each of the 5 video clips representing 5 different levels of skill:
Is the video representative?
Is there superfluous information in the video?

For each of the 3 audio clips :
Are the thoughts of the participants appropriate to the level of skill situation?
Is there superfluous information in the audio clips?

For each of the text sections - Bondy:
do the descriptions accurately reflect the level of skill being demonstrated?
Is the explanation sufficient? superfluous
For each of the text sections - Steinaker & Bell:

do the descriptions accurately reflect the level of skill being demonstrated?
Is the explanation sufficient? superfluous?

Overall impressions of this part of the program:

Levels of skill: self-assessment

Does the introductory text tell you what you think you need to know?
Is it obvious to you what is to be achieved in this part of the program?
Does this section achieve its aims?
Is the video clip essential?
Could the aims of this part of the program be achieved without the video clip?
How effective is this use of the visual medium?
Is the feedback adequate?
Is it essential that further details of errors be given in feedback?
Do you agree with the "correct" response?
Is there a sufficient number of video clips?
Does this section appeal to you?

Assessment Simulation
Does the introductory text tell you what you think you need to know?
Is it obvious to you what is to be achieved in this part of the program?
Do you think it will be obvious to mentors what is to be achieved here?
Is there anything unclear about the interface?
Are the icons consistent with the content they represent?
For the case studies generally,
Does the simulation mirror reality?
Are there any flaws in the approach to this section: philosophical, ethical theoretical, or otherwise?

For each case study:
Name of 'student".....

Is the introductory sentence from the student beneficial?
Is the portfolio of benefit to the user of the program?
Is the portfolio realistic?
Is the video too long? Long enough?
Does the video present sufficient material to assess the student?
Does the question and answer section contribute to the user's ability to judge the student's performance ?
Are all the domains of nursing practice identifiable on video?
Is it reasonable to ask of mentors to identify domains of practice and judge levels of skill simultaneously?
Are the video controls apparent? Easy to use?
Is the video sound quality a hindrance?
Is real-life video footage essential or would role-play be equally as useful?
Do the audio contributions of staff members add anything to the assessment of the student?
Is the assessment-entry grid intuitive? Counter-intuitive?

Feedback to user:
Is the analysis sufficiently accurate?
Does the feedback statement give sufficient direction?
What contribution does the "score" of your performance make?
Is the feedback, in total, satisfactory or not?

Summary:
would mentors apply themselves to using this, without help?
would this section motivate the mentors to explore other areas of the program in more depth?
General:
Which design features are most appealing and why?
Which design features are least appealing and why?

Summary
Overall Impressions of the Model
Overall Impressions of the Domains of Nursing Practice
Overall Impressions of the Levels of Skill
Overall Impressions of the Assessment Simulation
Overall Impressions of the Entire Program

- **Learning effectiveness**
- **Learner Interest/motivation**
- **Content quality**
- **Technical quality**
- **Implementability**

Appendix 16: NBS Conference

Mentorship in Nurse & Midwife Education

NBS

Wednesday 4th December 1996

The Hilton National Hotel, Edinburgh

The National Board for Nursing, Midwifery and Health Visiting for Scotland is organising a one-day conference in Edinburgh to publicise the findings of Professor Margot Cameron-Jones' project on **"The Role of the Nurse Mentor"**.

The target audience for this event will be nurse and midwife lecturers and service staff who fulfil the role of mentors. Five places will also be earmarked for students, for whom the registration fee will be waived and assistance with travelling expenses may be provided in appropriate cases. The number of places at this workshop will be restricted to a maximum of 80 and, should the event be oversubscribed, a waiting list will be operated by the organisers.

The programme will follow the format :

Registration and Coffee/Tea	**9.30 - 10.00**
Chairman's Introduction	
Professor Margaret Alexander	**10.00 - 10.15**
"The Role of the Nurse Mentor"	
Professor Margot Cameron-Jones	**10.15 - 11.00**
Coffee/Tea and Biscuits	**11.00 - 11.30**
"The Thorny Issue of Mentor, Preceptor or Clinical Supervisor"	
Mrs Morag Gray	**11.30 - 12.00**
Multi Media Presentation	
Mr Charles Docherty	**12.00 - 12.30**
Lunch	**12.30 - 13.30**
Concurrent Sessions - Choice of 1 from 3	**13.30 - 14.30**
Concurrent Sessions - As above	**14.30 - 15.30**
Coffee/Tea & Pastries	**15.30 - 15.50**
Open Forum	**15.50 - 16.15**
Summing Up by Chairman	**16.15 - 16.25**
Vote of Thanks and Close of Conference	**16.25 - 16.30**

There will be a registration fee of £35 payable to the National Board, and a Certificate of Attendance, which may be entered in participants' Personal Professional Profiles, will be awarded at the end of the event. The registration fee will not be refunded to individuals who cancel after the closing date. However, if participant numbers following the closing date do not justify proceeding with arrangements and the National Board takes the decision to cancel the event, the registration fee will be returned in full.

Further information may be obtained from **Dr Tricia Murphy-Black, Research and Developments Officer, NBS.**

Completed application forms, accompanied by the appropriate fee where applicable, should be returned by **FRIDAY 25th OCTOBER 1996** to :

Ms Liz Harden, Administrator, Research and Developments, NBS, 22 Queen Street, EDINBURGH, EH2 1NT

Appendix 17: NBS Conference - questionnaire

" Assessing Student Nurses" Multimedia Program

Questionnaire

Name
Position
Organisation

Having witnessed a short demonstration of the prototype version of this program, I would be grateful if you would spend some time advising me of its strengths and weaknesses to guide future development.

Thank you,

1. How well does this program achieve what it is designed to achieve?
   Please comment:

2. How interested would mentors be in using this program?
   Please comment:

3. How easy would it be to use programs such as this in your institution?
   Please comment:

4. Would you, in your institution, be interested in using this program if it was available to you?
   Yes **No** **(please tick)**
   Comments

5. Do you consider the *best* use of a program such as this to be...
   (please tick one)
   used as a workplace based resource?
   I n *addition to* conventional teaching in assessment?
   in *place of* conventional teaching in assessment ?
   comments:

6. Is the "model of the assessment process" part of the program similar to the assessment process in your institution?
   YES **NO** **(please tick)**
   If no, could you *briefly* describe your assessment strategy.

7. Are the concepts illustrated in the program those you would consider to be the most important in assessing student nurses' clinical performance?
(i.e., objectively **sampling the students performance** through Benner's Domains of nursing practice and **judging performance** through Bondy's criterion referenced scale?)

YES **NO** **(please tick)**

If no, could you please identify additional or alternative concepts that you think need to be expanded upon.

8. Does the assessment simulation part of the program ***mirror*** reality?

not at all
slightly
mostly
very much so

(please tick)

Comments:

9. Please comment on your overall impressions of the entire program in terms of -
a) Learning Effectiveness

b) Learner interest/motivation

c) Content quality

d) Technical quality

e) Implementability

f) Future developments

And finally,

10. Are there changes you could suggest which would make this program more viable as a general tool to assist mentors assess students?
Please comment:

Thank you for your time and views. Please return to me in the envelope provided.

Yours sincerely,

Charles Docherty.

Appendix 18: Collated formative evaluation recommendations

General

- Poor quality audio recordings should be re-digitised. If quality remains poor, then re-record using a DAT recorder. Similarly, existing video to be re-digitised and edited to remove unwanted background noise. If this is not possible to a reasonable standard, further filming should be considered. Continue to use video excerpts from real situations, as opposed to actors.
- Video controls should be explained to the user.
- The final program should be cross-platform compatible
- Replace 'go back' and 'quit' buttons with more accurate navigational cues
- Navigational and other instructions should generally be shown once and thereafter only made available on request.
- Reference to Glasgow College of Nursing and Midwifery should be removed.
- Backwards and forwards scrolling of text should be facilitated in all text screens.

Accompanying literature

- An accompanying instruction booklet with learning outcomes, references, technical information, and a 'quick guide' would be helpful. It should also contain advice as to how the program may be used.

Introduction and initial navigation instructions

- Timed screens in the introduction should be replaced.
- A 'guided tour' of the program should be offered as an option to first-time users. This would allow users to make informed choices as to where to begin. How to use hypertext links and video controls should also be incorporated into the guided tour. Audio could usefully reinforce text here.
- Introductory text and navigation instructions should be written in a less complex style.

Model of ' The Assessment Process'

- All reference to Glasgow College of Nursing and Midwifery should be removed .
- The model adjusted to reflect a more general assessment process.
- Reflection on practice to be enhanced and Link Tutor role to be less specific.
- Identify and edit cluttered text screens.
- Improve the 'model' diagram to achieve coherence in colour coding and remove ambiguity in hypertext links.

Concept Demonstration

Emphasis should be placed on the principles of adequacy in sampling, and consistency in judging students' performance in order that the program is more durable and more widely accepted.

- *Benner's Domains* should be renamed "Sampling Performance"
- *Levels of Skill* should be renamed "Judging Performance".

Domains of Nursing Practice ('Sampling Performance')

- An additional framework for sampling performance should be included in the concept demonstration part of the program.
- Self-assessment elements should give reasons as to why
- the correct answer *is* correct
- each of the incorrect choices *is* incorrect.
- Benner's photograph should be used in the introduction to her domains
- Each of Benner's Domains of nursing practice should include a practical example
- An Icon that betters represents 'sampling performance' should be developed.

Levels of skill ('Judging Performance')

- Self-assessment elements should give reasons as to why
- the correct answer *is* correct.
- each of the incorrect choices *is* incorrect.

- A choice of taxonomies should be available to help users identify different levels of skill.
- A sequence of accessing data in the 'levels of skill' selection screen should be recommended.

Simulation

- The process of data entry should be made explicit in the guided tour.
- Ambiguity in the assessment grid should be removed by having the numbers 1-5 hypertext linked to a definition of that particular level, for each of the taxonomies used.
- A choice of framework for sampling performance, in addition to Benner, should be included in the assessment simulation part of the program.
- The 'Question and Answer' section should be introduced in such a way that links are established between this part of the program and others.
- Hypothetical questions in the 'Question and Answer' part of the simulation should be consistent with other aspects of the simulation and should not be restricted in the number of questions that users can choose from.
- A statement explaining how the correct assessment is determined by the computer should be included at the beginning of the simulation.
- A sequence of accessing data in the simulation should be recommended.

Appendix 19: Topic Guide for lecturers' focussed interviews

Topic guide for focussed interviews

Do mentors worry about having to assess students?

Do you consider mentor anxiety to be something that affects how students are assessed?

Do mentors display different attitudes towards their role as mentor?

Do you consider mentor attitude to be something that affects how students are assessed?

Do mentors know and understand what is expected of them in their role as assessor?

How important is it for the mentor to know and fully understand their role?

Do you consider the mentor's knowledge and understanding to be something that affects how students are assessed?

Do you consider that mentors' self-efficacy has relevance to student assessment?

Are there other intrinsic influences on a mentor's ability to assess students?

Appendix 20: The Assessment Potential Questionnaire

Please note: this is the final version of the instrument, therefore does not contain 'knowledge' items.

Mentorship Questionnaire

Thank you for agreeing to participate in this exercise. The contents of this form are absolutely confidential. Your identity will not be disclosed under any circumstances.

Please provide the following information:

1 (a) **Name** (b) Male ❑
Female ❑

2 **Professional qualifications** *(please tick all that apply)*
RGN only ❑
Second Registered Qualification ❑
Nursing Degree or Diploma ❑

3 **Year of first registration**

4 **Clinical Grade**
G grade ❑
F grade ❑
E grade ❑
D grade ❑
other (please specify) ❑

5 **How involved have you been in the assessment of students?**
No involvement ❑
Some involvement ❑
Given overall responsibility for student assessment ❑

6 **Have you undertaken any of the following courses?**
(please tick any that apply)
ENB 998 Teaching and Assessing course ❑
SVQ Assessors course ❑
Professional Studies in Teaching, Counselling and Assessing ❑
Teaching certificate/diploma/degree ❑
Other similar (please state) ❑

7 **Which of these categories best describes your workplace?**
Specialised hospital unit, eg, ITU, A&E, Theatres ❑
Hospital ward ❑
Clinic or Day Hospital ❑

Appendix 20: The 'Assessment Potential Questionnaire' – page 2

Nursing Home or Care of the Elderly Hospital ❑
GP surgery or Health Centre ❑
Other (please state) ❑

8 **How often do you have students in your workplace?**

Almost constantly ❑
More often than not ❑
Occasionally ❑
Never ❑

Appendix 20: The 'Assessment Potential Questionnaire' – page 3

Mentor Self-evaluation

Read each statement and then tick the box to indicate how you feel *right now*, that is, *at this moment* as you consider your role in assessing student nurses.

There are no right or wrong answers. Do not spend too much time on any one statement, but give the answer which seems to describe your present feelings best.

		Not at all	Somewhat	Moderately	Very much
9	feel calm				
10	feel secure				
11	am tense				
12	feel strained				
13	feel at ease				
14	feel upset				
15	am presently worrying over possible misfortunes				
16	feel satisfied				
17	feel frightened				
18	feel comfortable				

Appendix 20: The 'Assessment Potential Questionnaire' – page 4

		Not at all	Somewhat	Moderately	Very much
19	feel self-confident				
20	feel nervous				
21	am jittery				
22	feel indecisive				
23	am relaxed				
24	am content				
25	am worried				
26	feel confused				
27	feel steady				
28	feel pleasant				

Appendix 20: The 'Assessment Potential Questionnaire' – page 5

Please read each statement and tick the appropriate box.

Make only one entry per statement.

		Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
29	It is enjoyable teaching students new skills					
30	Assessments require complete objectivity on the part of the mentor					
31	It is the mentor's fault if a student does not learn new skills					
32	Student assessments are high on my list of priorities					
33	Assessing students and giving them feedback can be enjoyable					
34	Students should accept their assessments without question					
35	Mentors should not criticise a student's performance					
36	Assessment is the best way of shaping a student's performance					
37	Students need to be constantly reminded of their weak points					
38	Assessing students is the most important mentorship function					
39	It is difficult to assess student nurses objectively					
40	Good students can instinctively be distinguished from bad.					

		Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
41	Assessing students is part and parcel of being a mentor					
42	I would not hesitate to fail a student whose performance was below the expected standard					
43	Assessing students is a process that has to be learned					
44	Failing an unsatisfactory student is worth any extra effort and stress that it may cause					
45	Students' self-assessments are important					
46	Good students just get on with it and take up little of my time					
47	Students should be encouraged to question decisions made regarding patient care					
48	Adequately assessing students demands too much of my time					
49	Mentors should support and encourage struggling students					
50	Student assessments are too demanding to complete properly					

AS OF NOW,

how *Confident* are you, that you are able to.....

		very confident	confident	some confidence	little confidence	no confidence whatsoever
51	maintain standards in the profession through your assessments of nursing students.					
52	fail a student without feeling guilty					
53	distinguish between those occasions when it is appropriate to assess students, from those when it is not					
54	assess a student without worrying about the consequences					
55	make a student feel at ease when being assessed					
56	separate the student's personality from their performance when making judgments					
57	judge a student objectively whether or not they respect you or your judgment					
58	judge success or failure by the criteria outlined in assessment documentation					
59	write an assessment that sums up all aspects of a student's abilities					
60	act as a role model for the student to base their performance upon					
61	distinguish between formative and summative assessment					

Appendix 20: The 'Assessment Potential Questionnaire' – page 8

		very confident	confident	some confidence	little confidence	no confidence whatsoever
62	assess a student within a minimum of 16 days on placement					
63	provide enough support for a student to be successful, who is initially under- performing					
64	with the student, identify items that are necessary to be included in their action plan					
65	use course assessment documents to determine what has to be assessed					
66	balance conflicting evidence when judging a student's performance					
67	identify different categories of skill through analysing overall student performance					
68	objectively assess different aspects of a student's performance					
69	write a critical assessment of the student's performance even though the student might be upset by it					
70	accurately assess a student in your own particular clinical situation					
71	identify types of evidence that may contribute to the assessment of a student's performance					
72	assess a student's performance without comparing the student to others					
73	state the criteria that determine whether or not an assessment is valid					

Thank you for participating .

Appendix 21: Mentors' letter: to accompany 'Assessment Potential Questionnaire'

GLASGOW

CALEDONIAN

UNIVERSITY

Stobhill Campus
557 3443
ext 207

Professor Barbara A Pavlin
PhD MSc MCommH ALBC RGN SCM RNT
Head of Department of
Nursing & Community Health

7 May 1998

Dear colleague,

At a mentorship study day/course recently it was explained to you that a research questionnaire is under development, and you kindly agreed to complete an early draft. Continuing with the development of this questionnaire, I would be obliged if you could fill one out a second time. As you are aware, this will take about 10 minutes. Please do this as soon as possible (today, please) and return to me in the envelope provided. Your assistance in this is greatly appreciated.

Thank you,

Charles Docherty
Nurse Teacher

WHO COLLABORATING CENTRE
FOR NURSING EDUCATION,
RESEARCH, PRACTICE & MANAGEMENT

City Campus
Cowcaddens Road
Glasgow G4 0BA
Telephone 0141 331 3000
Facsimile 0141 331 3005

Principal and Vice-Chancellor
Professor J S Mason
BSc PhD CEng FIMechE FIMarE MIMarE

Appendix 22: Protocol for telephone interviews

Focussed Telephone Interview Protocol No....

Hello I'm Charles Docherty of Glasgow Caledonian University undertaking research into mentorship preparation. I am interested in your views regarding your recent attendance at a mentorship half study day focusing on assessing student nurses' clinical performance. Can you spare three minutes answering a few short questions? Anything you have to say will be treated in strictest confidence.

7 How do you think that your recent study day has affected your ability to assess student nurses?

Helped Unchanged Confused Hindered

Why do you think this is?

2. Would you say the prospect of assessing students is now more or less likely to worry or trouble you?

More Less Unchanged

Why do you think this is?

3 How would you say your attitude towards students and education has been affected by attending the half study day?

Improved Unchanged Deteriorated

Would you like to comment further?

4 In conclusion, how would you say that your confidence in assessing students has been affected by attending the half study day?

Improved Unchanged Deteriorated

Would you like to say anything more ?

Thanks very much.

If there is anything further that you would like to discuss, please get back to me on 331 8337

Bye.

Appendix 23: Ethics committee approval letters for the Randomised Controlled Trial

Letter 1

GLASGOW CALEDONIAN UNIVERSITY

REF: BAP/AM/24

DATE: 19 October 1998

FROM: Professor Barbara A Parfitt, HoD, NCH

TO: Mr Charles Docherty, NCH

Thank you for your note regarding your planned randomised controlled trial to evaluate the multimedia interactive learning program 'A Mentor's Guide to Assessing Student Nurses'.

I give my consent for you to proceed to NCH Ethics Committee with your proposal.

Barbara

BARBARA A PARFITT

Letter 2

GLASGOW CALEDONIAN UNIVERSITY

DT/mb

09 November 1998

Mr Charles Docherty
Department of Nursing and Community Health
Glasgow Caledonian University
City Campus
GLASGOW
G4 0BA

Professor Barbara A Parlin
PhD MSc MCommH ALBC RGN SCM RNT FN
Head of Department of
Nursing & Community Health

Dear Mr Docherty

RE: Application for ethical approval of proposed study:
"The Design and evaluation of a multimedia program to improve mentors' assessments of student nurses"

On behalf of the Departmental Ethics Committee, I am pleased to approve your proposed study and wish you success in the completion of this investigation.

Yours sincerely

DR. DEBBIE TOLSON, CHAIR
DEPARTMENTAL ETHICS COMMITTEE

WHO COLLABORATING CENTRE
FOR NURSING AND MIDWIFERY
EDUCATION, RESEARCH AND PRACTICE

City Campus
Cowcaddens Road
Glasgow G4 0BA
Telephone 0141 331 8311
Facsimile 0141 331 8312
E-Mail B.A.Parlin@gcal.ac.uk

Appendix 24: Poster - mentorship study days.

Attention – New Mentors!

Clinical Assessment Workshops

Workshops have been designed to introduce new mentors to the principles and practice of assessing student nurses. These will be held in the new Health Building of Glasgow Caledonian University.

Who should attend?
These workshops are for staff nurses new to mentorship who have not been formally prepared for involved in student nurse assessment. Each of these workshops is identical and can be used as an alternative to the 'assessment' half study days in existing programmes of on-going support already established throughout Glasgow. Attendance will be recognised with a certificate that may be used for PREP.

Please note: *by attending these particular workshops you will be asked to complete two questionnaires and may be expected to use a computer simulation program. This is part of a research project approved by Glasgow Caledonian University and supervised by Glasgow University.*
Should you not wish to participate in this research, please be reassured that conventional workshops will continue to be provided as part of the programme of on-going mentor support already established in your geographical area.

Location
Glasgow Caledonian's new Health Building.

Date	Time	Room
28 January	**1.30pm**	**A427**
18 February	**1.30pm**	**A427**
11 March	**1.30pm**	**A427**
8 April	**1.30pm**	**A427**
29 April	**1.30pm**	**A427**

Registration
For more information and registration, please contact Charles Docherty directly:
Telephone: 0141 331 8337 or 01698 333 710
email: 7423221d@student.gla.ac.uk or C.A.Docherty@gcal.ac.uk
Fax 0141 331 8312

Places are limited and will be offered on a first-come first-served basis. Confirmation of registration will be made at the time of your call.

Appendix 25: Guidelines on using the program during the RCT

Instructions on Getting Started

Click mouse

.....a box appears asking for identity and password.

- **Look** at the white sign above the screen
- **Type** in the first line, the code that looks like:

g3-1

...this will be different for each machine, so use only the code that is on your computer.

- **Using the mouse**, move the cursor to the second line, and type the same code again.

5 **Double-click** on the 'hard disc' on the top right hand corner of the screen.

- You should now see a variety of other objects: look for a box called **'Mentor'** double-click on this box.
    - Again, double click on a similar looking box.

Once inside the program,

Click "Guided Tour"

After the tour

- LOG-On
- Go to "SIMULATION"
- Select "Tracey"
- work on this, initially, then you're on your own to explore.

Appendix 26: Content & Context of the teaching of 'Assessment' to the control groups

The traditional teaching of 'Assessment' that acted as control within the RCT was one of a series of half study days that served the purpose of preparing mentors to undertake their role.

The other subjects taught as part of this preparation varied slightly between different zones within Glasgow. Subjects within one zone, the 'Victoria Infirmary NHS Trust and Independent Sector', in 1999, included

- Mentorship concept and roles
- Student nurses: learners or workers?
- Current courses of preparation for registration
- Giving and receiving feedback
- The Assessment Process
- Developing and using Action Plans
- Teaching Nursing Skills
- Mentor – student relationships.

Each session was self-contained, and lasted two hours, teaching methods included information-giving, discussion and problem solving. For 'The Assessment Process',
The following learning outcomes were typical:

Following the session, the prospective mentor should

- Have a knowledge of the extent and component parts of the assessment process
- Understand the purpose of clinical assessment
- Be able to categorise different aspects of a student's clinical practice
- Be able to apply criteria to the judging of clinical performance
- Be able to apply theoretical concepts to the practice of clinical assessments
- Understand the principles underpinning how the assessment document is constructed and how it should be completed.

References

1. Shortell, S., and Richardson, W. C., Health Program Evaluation. C.V.Mosby, St. Louis, 1978.

2. Polit, D. F., and Hungler, B., Nursing Research Principles and Methods. Lippincott, Philadelphia, 1995.

3. UKCC, The Council's Proposed Standards for Post-registration Education. United Kingdom Central Council for Nursing, Midwifery and Health Visiting, London, 1993.

4. UKCC, P2000: A new preparation for practice. United Kingdom Central Council for Nursing, Midwifery and Health Visiting. London, 1986.

5. UKCC, The Council's proposed Standards for Education and Practice Following Registration: Programmes of Education leading to the Qualification of Specialist Practitioner. United Kingdom Central Council for Nursing Midwifery and Health Visiting, London, 1994.

6. Schon, D. A., The Reflective Practitioner. Temple Smith, London, 1983.

7. Bennett, Y., The validity and reliability of assessments and self assessments of work - based learning. Assessment and Evaluation in Higher Education, 18(2) 83-94 (1993).

8. Watkins, M. J., Competency for nursing practice. Journal of Clinical Nursing, 9(3) 338-346 (2000).

9. Girot, E. A., Assessment of graduates and diplomates in practice in the UK - are we measuring the same level of competence? Journal of Clinical Nursing, 9(3) 330-336 (2000).

10. Erault, M., and Cole, G., The assessment of competence in the professions. University of Sussex, London, 1992.

11. NBS, Partnerships in Development and Delivery. National Board for Nursing Midwifery and Health Visiting for Scotland, Edinburgh, 2000.

12. May, N., et al., Preparation for Practice: Evaluation of Nurse & Midwife Education in Scotland. Glasgow Caledonian University, Glasgow, 1997.

13. Watson, H., and Harris, B., Supporting Students in Practice Placements in Scotland. Glasgow Caledonian University for the National Board for Nursing, Midwifery and Health Visiting for Scotland., Glasgow, 1999.

14. Duffy, K., et al. The nurse lecturer's role in mentoring the mentors. Nursing Standard, 15(6) 35-38 (2000).

15. Neary, M., Phillips, R., and Davies, B., The introduction of Project 2000 in Wales. Nursing Standard, 10(25) 37-39 (1996).

16. Wilson-Barnett, J., et al., Clinical support and the project 2000 nursing student: factors influencing this process. Journal of Advanced Nursing, 21 1152-1158 (1995).

17. Twinn, S., and Davies, S., The supervision of Project 2000 students in the clinical setting: issues and implications for practitioners. Journal of Clinical Nursing, 5 177-183 (1996).

18. White, E., et al., A detailed study of the relationships between teaching, support, supervision and role modelling for students in clinical areas, within the context of Project 2000 courses. ENB, London, 1993.

19. Richardson, L., The medium and the message. Australian Journal of Educational Technology, 11(1) http://cleo.murdoch.edu.au/gen/aset/ajet11/wi95p1.html, 1-11 (1995).

20. Dearing, R., Higher Education in the Learning Society: Report of the National Committee of Inquiry into Higher Education. HMSO, London, 1997.

21. Garrick, R., Higher Education in the learning Society. Report of the Scottish Committee. HMSO, Edinburgh, 1997.

22. QAA, Guidelines on the Quality Assurance of Distance Learning, Gloucester, 1999.

23. McNay, I., and Bailley, B., Education Policy Report. University of Greenwich, London, 2000.

24. Hall, W., As we may learn: the application of hypermedia in education. Computing and Control Engineering Journal, 11(4) 173-178 (2000).

25. RCN, The Nursing Reconstruction Committee (Horder Report). HMSO, London, 1943.

26. Briggs, A., Report of the Committee on Nursing (Briggs Report). HMSO, London, 1972.

27. Clark, J. M., Maben, J., and Jones, K., Project 2000. Perceptions of the philosophy and practice of nursing: preparation for practice. Journal of Advanced Nursing, 26(2) 246-256 (1997).

28. May, N., and Domokos, T., Interim Evaluation Report 1:Issues in the early development of Project 2000. Glasgow Caledonian University, Glasgow, 1993.

29. Elkan, R., and Robinson, J., Project 2000: the gap between theory and practice. Nurse Education Today, 13 295-298 (1993).

30. Elkan, R., and Robinson, J., Project 2000: a review of published research. Journal of Advanced Nursing, 22 386-392 (1995).

31. Parker, T. J., and Carlisle, C., Project 2000 students' perceptions of their training. Journal of Advanced Nursing, 24 771-778 (1996).

32. Phillips, T., et al., Assessment of competencies in nursing and midwifery education and training (the ACE project). Research Highlights No. 4. ENB, London, 1993.

33. Jowett, S., Walton, I., and Payne, S., Implementing Project 2000-an interim report. NFER, Slough, 1992.

34. French, H. P., Educating the nurse practitioner: an assessment of the pre-registration preparation of nurses as an educational experience. University of Durham, Durham, England, 1990.

35. DOH, Project 2000: Fitness for purpose. Universities of Warwick and Liverpool, London, 1996.

36. Runciman, P., Dewar, B., and Goulbourne, A., Employer's needs and the skills of newly qualified Project 2000 staff nurses. NBS, Edinburgh, 1998.

37. Luker, K., et al., Project 2000 Fitness for purpose: Report to the Department of Health. University of Warwick and University of Liverpool, London, 1996.

38. Carlisle, C., et al., Skills competency in nurse education: nurse managers' perceptions of diploma level preparation. Journal of Advanced Nursing, 29(5) 1256-1264 (1999).

39. Peach, L., Fitness for Practice, the UKCC Commission for Nursing and Midwifery Education. United Kingdom Central Council for Nursing, Midwifery and Health Visiting, London, 1999.

40. Joyce-Nagata, B., Reeb, R., and Burch, S., Comparison of Expected and Evidenced Baccalaureate Degree Competencies. Journal of Nursing Education, 28(7) 314-321 (1989).

41. Gerrish, K. B., Still fumbling along? A comparative study of the newly qualified nurse's perception of the transition from student to qualified nurse. Journal of Advanced Nursing, 32(2) 473-480 (2000).

42. Kieffer, J. S., Selecting Technical Skills to Teach for Competency. Journal of Nursing Education, 23(5) 198-203 (1984).

43. Rowe, H., Managing NVQ's in the workplace. Nursing Management, 1(5) 27-28 (1994).

44. Storey, L., O'Kell, S., and Day, M., Utilising National Occupational Standards as a complement to nursing curricula. Department of Health, London, 1995.

45. Bradshaw, A., Defining competency in nursing (part II): an analytical review. Journal of Clinical Nursing, 7(2) 103-111 (1998).

46. Kevern, J., Ricketts, C., and Webb, C., Pre-registration diploma students: a quantitative study of entry characteristics and course outcomes. Journal of Advanced Nursing, 30(4) 785-795 (1999).

47. Newton, G., Talking up nursing. Nursing Standard, 11(2) 16-17 (1996).

48. Norman, I., et al., Evaluation of the validity and reliability of methods to assess the competence to practise of pre-registration nursing and midwifery students in Scotland. NBS, Edinburgh, 2000.

49. Wood, V., Clinical evaluation of student nurses: syllabus needs for nursing instructors. Nurse Education Today, 6 208-214 (1986).

50. Woolley, A. S., The Long and Tortured History of Clinical Evaluation. Nursing Outlook, 25(5) 308-315 (1977).

51. While, A. E., The Problem of Clinical Evaluation - a review. Nurse Education Today, 11 448-453 (1991).

52. Ramsborg, G. C., Evaluation of clinical performance, part 1. Journal of the American Association of Nurse Anesthetists, 51(1) 55-62 (1983).

53. Lankshear, A., Failure to Fail: the Teacher's Dilemma. Nursing Standard, 4(20) 35-37 (1990).

54. Krichbaum, K., et al., The Clinical Evaluation Tool: A Measure of the Quality of Clinical Performance of Baccalaureate Nursing Students. Journal of Nursing Education, 33(9) 395-404 (1994).

55. Rowntree, D., Assessing Students: How Shall We Know Them? Kogan Page, London, 1987.

56. Burnard, P., The Journal as an Assessment and Evaluation Tool in Nurse Education. Nurse Education Today, 8 105-107 (1988).

57. Andrusyszyn, M. A., Clinical evaluation of the affective domain. Nurse Education Today, 9 75-81 (1989).

58. Alspach, G., Concern and Confusion over Competence. Critical Care Nurse, 12(4) 9-11 (1992).

59. Burrows, E., Clinical Practice: an approach to the assessment of clinical competencies. British Journal of Occupational Therapy, 52(6) 222-227 (1989).

60. Foster, S., and Biddle, C. J., Performance Contracts in Clinical Evaluation. Journal of the American Association of Nurse Anesthetists, 53(6) 497-503 (1985).

61. Purdy, M., The problem of self-assessment in nurse education. Nurse Education Today, 17 135-139 (1997).

62. Dunn, D. M., Have you Done my Report Please Sister? Nursing Times Quest, 4(2) 56-59 (1984).

63. Hepworth, S., The Assessment of Student Nurses. Nurse Education Today, 11 46-52 (1991).

64. Yuen, F., Pincombe, and Martin, M., Student Competence Profile: The Wollongong Approach. International Nursing Review, 34(4) 107-109 (1987).

65. Weller, L. D., The Grading Nemesis: an Historical Overview and a Current Look at Pass/Fail Grading. Journal of Research and Development in Education, 17(1) 39-45 (1983).

66. Gerrish, K., McManus, M., and Ashworth, P., Levels of achievement: a review of the assessment of practice. English National Board, London, 1997.

67. Ashworth, P., et al.,'Levels' of attainment in nursing practice: reality or illusion? Journal of Advanced Nursing, 30(1) 159-168 (1999).

68. Day, M., NVQ Assessments for Care Assistants. Nursing Standard, 7(29) 35-38 (1993).

69. Ridley, J., Laschinger, H. K., and Goldenberg, D., The effect of a senior preceptorship on the adaptive competencies of community college nursing students. Journal of Advanced Nursing, 22 58-65 (1995).

70. Morle, K. M. F., The problems in evaluating student nurses. Nurse Education Today, 4(2) 80-85 (1984).

71. SOHHD, The role and function of the professional nurse. The Scottish Office National Nursing and Midwifery Consultative Committee., Edinburgh, 1992.

72. Chambers, M., Some issues in the assessment of clinical practice: a review of the literature. Journal of Clinical Nursing, 7 201-208 (1998).

73. Squier, R. W., The reliability and validity of rating scales in assessing the clinical progress of psychiatric nursing students. International Journal of Nursing Studies, 18(3) 157-169 (1981).

74. Chandler, J., Reforming nurse education 2 - implications for teachers and students. Nurse Education Today, 11 89-93 (1991).

75. Crooks, T. J., The impact of classroom evaluation practices on students. Review of Educational Research, 58(4) 438-481 (1988).

76. Ecclestone, C., Understanding Assessment: a guide for teachers and managers in post-compulsory education. National Institute of Adult Continuing Education, Leicester, 1994.

77. Alavi, C., Problem-based Learning in a Health Sciences Curriculum. Routledge, London, 1995.

78. Boyd, W., and King, E. J., The History of Western Education. Black, London, 1980.

79. Jarvis, P., Adult and Continuing Education. Routledge, London, 1995.

80. Eraut, M., Developing Professional Knowledge and Competence. Falmer Press, London, 1994.

81. Burnard, P., Learning from experience: nurse tutors' and student nurses' perception of experiential learning in nurse education: some initial findings. International Journal of Nursing Studies, 29(2) 151-161 (1992).

82. Boud, G., and Feletti, G. E., The Challenge of Problem Based Learning. Kogan Page, London, 1997, pp. 344.

83. Smith, L., Experiential Learning. *In* Curriculum planning in nursing education, edited by S. Pendleton and A. Myles. Edward Arnold, London, 1991, pp. 144-165.

84. Maynard, C. A., Relationship of critical thinking ability to Professional Nursing Competence. Journal of Nursing Education, 35(1) 12-18 (1996).

85. Glen, S., Developing critical thinking in higher education. Nurse Education Today, 15 170-176 (1995).

86. Girot, E. A., Preparing the practitioner for advanced academic study: the development of critical thinking. Journal of Advanced Nursing, 21 387-394 (1995).

87. Dobrzykowski, T. M., Teaching Strategies to Promote Critical Thinking Skills in Nursing Staff. Journal of Continuing Education in Nursing, 25(6) 272-276 (1994).

88. Benner, P., From Novice to Expert: Excellence and Power in Clinical Nursing Practice. Addison- Wesley, Menlo Park, CA, 1984.

89. Benner, P., Tanner, C., and Chesla, C., From Beginner to expert: Gaining a Differentiated Clinical World in Critical Care Nursing. Advances in Nursing Science, 14(3) 13-28 (1992).

90. Benner, P., From Novice To Expert. American Journal of Nursing, 82(3) 402-407 (1982).

91. English, I., Intuition as a Function of the Expert Nurse: a Critique of Benner's Novice to Expert Model. Journal of Advanced Nursing, 18 387-393 (1993).

92. Sharp, K. J., et al., A literature review on competence to practise. National Board for Nursing, Midwifery and Health Visiting for Scotland., Edinburgh, 1995.

93. Cash, K., Benner and expertise in nursing: a critique. International Journal of Nursing Studies, 32(6) 527-534 (1995).

94. Darbyshire, P., Skilled Expert Practice: is it 'all in the mind'? A Response to English's critique of Benner's novice to expert model. Journal of Advanced Nursing, 19 755-761 (1994).

95. Rudge, T., Reflections on Benner: a Critical Perspective. Contemporary Nurse, 1(2) 84-89 (1992).

96. Carlson, L., Crawford, N., and Contrades, S., Nursing Student Novice to Expert - Benner's Research applied to Education. Journal of Nursing Education, 28(4) 188-190 (1989).

97. Darbyshire, P., et al., New Domains in Nursing. Nursing Times, 86(27) 73-75 (1990).

98. Darbyshire, P., In defence of pedagogy: a critique of the notion of andragogy. Nurse Education Today, 13 328-335 (1993).

99. Gatley, E. P., From Novice to Expert: the use of intuitive knowledge as a basis for district nurse education. Nurse Education Today, 12 81-87 (1992).

100. Bevis, E. O., Accessing Learning: determining worth or developing excellence - from a behaviourist toward an interpretive-criticism model. *In* Toward a caring curriculum: a new pedagogy for nursing. National League for Nursing, New York, 1989.

101. Burnard, P., Experiential learning and andragogy - negotiated learning in nurse education: a critical appraisal. Nurse Education Today, 9 300-306 (1989).

102. Kolb, D. A., Experiential Learning: Experience as the Source of Learning and Development. Prentice Hall, New Jersey, 1984.

103. Piaget, J., The origin of intelligence in the child. Routledge & Kegan, Paul, London, 1953.

104. Erikson, E. H., Childhood and Society. Penguin, Hamondsworth, 1965.

105. Burnard, P., Experiential Learning: some theoretical considerations. International Journal of Lifelong Education, 7(2) 127-133 (1988).

106. Steinaker, N. W., and Bell, R. M., The Experiential Taxonomy: a New Approach to Teaching and Learning. Academic Press, London, 1979.

107. Bloom, B. S., Taxonomy of educational objectives: Handbook I: Cognitive Domain. McKay, New York, 1956.

108. Sheahan, J., Measurement in Nursing Education. Journal of Advanced Nursing, 4 47-56 (1979).

109. Brooke, P., Nyatanga, L., and Walker, I., Facilitating the Self-Care Concept Using The Experiential Taxonomy. Senior Nurse, 9(2) 10-11 (1989).

110. Faltermeyer, T., The Higher Award - Assessing at the Differing Levels. Senior Nurse, 13(7) 32-36 (1994).

111. Nyatanga, L., The E T Penumbra. Senior Nurse, 10(7) 12-14 (1990).

112. Jones, E. E., and Nisbett, R. E., The actor and the observer: divergent perceptions of the causes of behaviour. *In* Attributions, edited by E. E. Jones. General Press, New Jersey, 1972.

113. Orchard, C., The nurse educator and the nursing student: a review of the issue of clinical evaluation procedures. Journal of Nursing Education, 33(6) 245-251 (1994).

114. Winter, R., Work-based Learning and Quality Assurance in Higher Education. Assessment and Evaluation in Higher Education, 19(3) 247-257 (1994).

115. UKCC, Nursing and Midwifery Competencies. Register, (31) 6-7 (2000).

116. Marriott, A., The support, supervision and instruction of nurse learners in clinical areas: a literature review. Nurse Education Today, 11 261-269 (1991).

117. Adderly, B. V., and Brock, A. M., Evaluating clinical performance in nursing. Journal of Advanced Nursing, 2 355-363 (1977).

118. Bashook, P. G., Clinical Assessment: A State of the Art Review. The Diabetes Educator, 11(supplement) 30-36 (1985).

119. Mahara, M. S., A perspective on clinical evaluation in nursing education. Journal of Advanced Nursing, 28(6) 1339-1346 (1998).

120. Pavlish, C., A model for clinical performance evaluation. Journal of Nursing Education, 26(8) 338-339 (1987).

121. Blomquist, K. B., Evaluation of students: intuition is important. Nurse Educator, 10(6) 8-11 (1985).

122. Hager, P., Gonczi, A., and Athanasou, J., General issues about assessment of competence. Assessment and Evaluation in Higher Education, 19(1) 3-16 (1994).

123. McGaghie, W. C., The evaluation of competence: validity issues in the health professions. Evaluation and the Health Professions, 3(3) 289-320 (1980).

124. Ashworth, P., and Morrison, P., Problems of competence-based nurse education. Nurse Education Today, 11 256-260 (1991).

125. Manpower Services Commission, Definition of Competence and Performance Criteria, the Development of Assessable Standards for National Certification. Manpower Services Commission, London, 1998.

126. Gardner, R., Competent to Nurse? Senior Nurse, 7(5) 3 (1987).

127. Girot, E. A., Assessment of competence in clinical practice - a review of the literature. Nurse Education Today, 13 83-90 (1993).

128. Umiker, W., The challenge of competency assessment. Health Care Supervisor, 17(3) 11-17 (1999).

129. McCloskey, J. C., Nursing Education and Job Effectiveness. Nursing Research, 32(1) 53-58 (1983).

130. Watson, A. B., Professional socialisation of the registered nurse. *In* Review of research in nursing education, edited by W. L. Holzemer. Thorofare., New Jersy, 1983.

131. Green, G. J., Relationships between role models and role perceptions of new graduate nurses. Nursing Research, 37(4) 245-248 (1988).

132. Ellis, R., Competence in the Caring Professions. *In* Professional Competence and Quality Assurance in the Caring Professions, edited by R. Ellis. Croom Helm, London, 1988, pp. 43-57.

133. While, A. E., Competence versus Performance: which is more important? Journal of Advanced Nursing, 20 525-531 (1994).

134. Rethans, J. J., et al., Competence and Performance: two different concepts in the assessment of quality of medical care. Family Practice, 7(3) 168-174 (1990).

135. Messick, S., The Psychology of Educational Measurement. Journal Of Educational Measurement, 21 215-238 (1984).

136. Medley, D. M., Teacher competency testing and the teacher educator. *In* Advances in Teacher Education, vol. 1, edited by L. G. Katz and J. D. Raths. Ablex Norwood, New Jersey, 1984, pp. 51-94.

137. Lloyds, J. S., Definitions of Competence in Specialities of Medicine. American Board of Medical Specialities, Chicago, 1979.

138. DeMers, J. L., Observational Assessment of Performance. *In* Evaluating Clinical Performance in Health Professionals, edited by M. Morgan and D. Irby. Mosby, St. Louis, 1978, pp. 89-105.

139. White, E., A scheme for continuous practical assessment. Senior Nurse, 10(3) 27-29 (1989).

140. Flagler, S., Loper-Powers, S., and Spitzer, A., Clinical teaching is more than evaluation alone! Journal of Nursing Education, 27(8) 342-348 (1988).

141. Ross, M., et al., Using The OSCE to measure clinical skills performance in nursing. Journal of Advanced Nursing, 13 45-56 (1988).

142. Johnson, G., Lehman, B. B., and Sandoval, J. B., Clinical Exam: A Summative Evaluation Tool. Journal of Nursing Education, 27(8) 373-374 (1988).

143. UKCC, The Nurses Midwives and Health Visitors Rules Approval Order. HMSO, (1983).

144. Whittington, D., and Boore, J., Competence in Nursing. *In* Professional Competence and Quality Assurance in the Caring Professions, vol. 1, edited by R. Ellis. Croom Helm, London, 1988, pp. 109 - 139.

145. Vaughan, B., Writing for competencies. *In* Open Distance Learning for Nurses, edited by K. Robinson. Longman, London, 1990.

146. Ellis, R., Professional competence and quality assurance in the caring professions. Croom Helm, London, 1988.

147. Turrell, S., Phillips, D., and Attrill, V., Competency Scoring. Nursing Times, 86(10) 61-62 (1990).

148. Chapman, H., Some important limitations of competency-based education with respect to nurse education: an Australian perspective. Nurse Education Today, 19(2) 129-135 (1999).

149. Runciman, P., et al., The work of newly qualified staff nurses: perceptions of core skills. Queen Margaret University College, Edinburgh, 2000.

150. Benner, P., Issues in Competency Based Testing. Nursing Outlook, 303-309 (1982).

151. Taylor, J., and Taylor, D., The assessment of vocational training in general medical practice. University of York, York, 1988.

152. Runciman, P., Competence Based Education and the Assessment and Accreditation of Workbased Learning in the Context of Project 2000 Programmes of Nurse Education: A literature review. The National Board for Nursing, Midwifery and Health Visiting for Scotland., Edinburgh, 1990.

153. Tomey, A. M., and Alligood, M. R., Nursing Theorists and their Work. Mosby, St Louis, 1998.

154. Gilmore, A., Report of the analysis of the literature evaluating pre-registration nursing and midwifery education in the UK. UKCC, London, 1999.

155. Debling, G., When the assessing has to stop. *In* Competence and Assessment Compendium, edited by E. Fennell. Employment Department, Sheffield, 1992, pp. 39.

156. Popham, J., Specifying the domain of content or behaviours. *In* A guide to criterion referenced test construction, edited by R. A. Berk. Johns Hopkins Press, Baltimore, 1984.

157. Le Var, R. M. H., NVQ's in nursing, midwifery and health visiting: a question of assessment and learning? Nurse Education Today, 16 85-93 (1996).

158. Swanwick, M., Observation as a research method. Nurse Researcher, 2(2) 4-12 (1994).

159. Johnson, D. M., and Wilhite, M. J., Reliability and Validity of Subjective Evaluation of Baccalaureate Program Nursing Students. Nursing Research, 22(3) 257-262 (1973).

160. Watson, H., Psychometric analysis of assessment tools. Nurse Researcher, 2(4) 56-68 (1995).

161. Newble, D. I., Hoare, J., and Elmslie, R. G., The Validity and Reliability of a New Examination of the Clinical Competence of Medical Students. Medical Education, 15 46-52 (1981).

162. Rudolph, E., Kaiser, K. A., and Corrigan, M. J., Construct Validity of a Clinical Performance Evaluation Tool in Community Health Nursing. Public Health Nursing, 4(2) 89-98 (1987).

163. Coates, V. E., and Chambers, M., Evaluation of Tools to Assess Clinical Competence. Nurse Education Today, 12 122-129 (1992).

164. Behi, R., and Nolan, M., Reliability: consistency and accuracy in measurement. British Journal of Nursing, 4(8) 472-475 (1995).

165. Gibbon, B., Validity and Reliability of assessment tools. Nurse Researcher, 2(4) 48-55 (1995).

166. Guilford, J. P., Psychometric Methods. McGraw Hill, New York, 1974.

167. Bondy, K. N., Clinical Evaluation of Student Performance: The Effects of Criteria on Accuracy and Reliability. Research in Nursing and Health, 7(1) 25-33 (1984).

168. Litwin, M. S., How to measure survey reliability and validity. Sage Publications, London, 1995.

169. Bateson, N., Data construction in Social Surveys. Allen and Unwin, London, 1984.

170. Nolan, M., and Behi, R., Validity: a concept at the heart of research. British Journal of Nursing, 4(9) 530-533 (1995).

171. Department, E., Competence and assessment briefing series No. 9. Employment Department Learning Methods Branch, Sheffield, 1993.

172. Schwirian, P. M., Evaluation research in computer based instruction. Computers in Nursing, 5(4) 128-131 (1987).

173. Bandura, A., Self-efficacy: toward a unifying theory of behavioural change. Psychol Rev, 2 191-215 (1977).

174. Jinks, A. M., Conceptualization of differing levels of educational attainment: what are the characteristics of nurses and midwives who have undertaken diploma and degree educational programmes? Journal of Nursing Management, 2 279-285 (1994).

175. Fitzpatrick, J. M., While, A. E., and Roberts, J. D., The measurement of nurse performance and its differentiation by course of preparation. Journal of Advanced Nursing, 20 761-768 (1994).

176. Giger, J. N., and Davidhizar, R. E., Conceptual and theoretical approaches to patient care: associate versus baccalaureate degree prepared nurses. Journal of Advanced Nursing, 15 1009-1015 (1990).

177. Ryan, M. E., and Hodson, K. E., Employer evaluations of nurse graduates: a critical programme assessment element. Journal of Nursing Education, 31 198-202 (1992).

178. Bartlett, H., et al., An evaluation of pre-registration nursing education: a literature review and comparative study of graduate outcomes. Oxford Centre for Health Care Research and Development. Oxford Brookes University, Oxford., Oxford, 1998.

179. Greenwood, J., and King, M., Some surprising similarities in the clinical reasoning of 'expert' and 'novice' orthopaedic nurses: report of a study using verbal protocol analyses. Journal of Advanced Nursing, 22 907-913 (1995).

180. Marquis, B., The Relationship Among Multiple Assessments of Nursing Education Outcomes. Journal of Nursing Education, 31(1) 33-38 (1992).

181. Schwirrian, P., and Gortner, S., How Nursing Schools Predict their successful graduates. Nursing Outlook, 27 352-358 (1979).

182. Lloyd Jones, R., and Bray, E., Assessment from Principles to Action. Macmillan, London, 1986.

183. Dunn, D. M., Assessing the Development of Clinical Nursing Skills. Nurse Education Today, 6 28-35 (1986).

184. Stecchi, J. M., et al., Comprehensive approach to clinical evaluation: one teaching team's solution to clinical evaluation of students in multiple settings. Journal of Nursing Education, 22(1) 38-46 (1983).

185. Cottrell, B. H., et al., A clinical evaluation tool for nursing students based on the nursing process. Journal of Nursing Education, 25(7) 270-274 (1986).

186. Dreyfus, H. I., and Dreyfus, S. E., Mind Over Machine. MacMillan, The Free Press, New York, 1986.

187. Fullerton, J. T., Piper, D. M., and Hunter, L., Development of a single component of a performance evaluation protocol - the faculty impression score. Journal of Nurse-Midwifery, 38(4) 236-240 (1993).

188. Hillegas, K. B., and Valentine, S., Development and evaluation of a summative clinical grading tool. Journal of Nursing Education, 25(5) 218-220 (1986).

189. Gennaro, S., et al., The Birth, Life, and Times of a Clinical Evaluation Tool. Nurse Educator, 1 27-32 (1982).

190. Wysocki, R., Evaluation of Student Clinical Performance. The Australian Nurses' Journal, 10(5) 42-43 (1980).

191. Tower, B. L., and Majewski, T. V., Behaviourally Based Clinical Evaluation. Journal of Nursing Education, 26(3) 120-123 (1987).

192. Aggleton, P., Allen, M., and Montgomery, S., Developing a System for the Continuous Assessment of Practical Nursing Skills. Nurse Education Today, 7 158-164 (1987).

193. Eaton, L. A., Clinical evaluation tool that gives students feedback. Nurse Educator, 20(3) 9-10 (1995).

194. Shearer, M., Progressing Nicely. Nursing Times, 85(3) 60-61 (1989).

195. Schwirian, P. M., Evaluating the performance of Nurses: a Multidimensional Approach. Nursing Research, 27(6) 347-351 (1978).

196. Bondy, K. N., Criterion Referenced Definitions for Rating Scales in Clinical Evaluation. Journal of Nursing Education, 22(9) 376-382 (1983).

197. Rheault, W., and Coulson, E., Use of the Rasch Model in the Development of a Clinical Competence Scale. Journal of Physical Therapy Education, 5(1) 10-13 (1991).

198. Springer, P., Payne, K., and Petermann, B., Rating Nursing Performance Based on Behaviours. Nurse Educator, 28(1) 39-45 (1998).

199. Krumme, U. S., The Case For Criterion Referenced Measurement. Nursing Outlook, 23(12) 764-770 (1975).

200. Fullerton, J. T., Greener, D. L., and Gross, L. J., Criterion Referenced Competency Assessment and the National Certification Examination in Nurse-Midwifery. Journal of Nurse-Midwifery, 34(2) 71-74 (1989).

201. Hawly, R., and Lee, J., Standardised Clinical Evaluation using Bondy Rating Scale. Australian Journal of Advanced Nursing, 8(3) 6-10 (1991).

202. Donoghue, J., and Pelletier, S. D., An Empirical Analysis of a Clinical Assessment Tool. Nurse Education Today, 11 354-362 (1991).

203. Cox, J. M., Bottoms, R. J., and Ramsey, J., Assessment of practice in pre-registrationnurse education and the development of a skills acquisition manual. Nurse Education Today, 18 190-198 (1998).

204. Underwood, I. M., and Reed, S. E., A strategy for continuous assessment of theory and practice in first level nursing courses. Nurse Education Today, 10 307-317 (1990).

205. Wilkinson, J., A practical guide to assessing students in clinical practice. British Journal of Nursing, 8(4) 218-222 (1999).

206. SCOTVEC, Notes on Assessment for Assessors of Scottish Vocational Qualifications in Care. Scottish Vocational Education Publication., Glasgow, 1992.

207. RCN, Report of a working party on ward-based assessments. RCN, London, 1978.

208. Neary, M., Preparing assessors for continuous assessment. Nursing Standard, 13(18) 41-47 (1999).

209. Jinks, A. M., and Williams, R. H., Evaluation of a community staff preparation strategy for the teaching, assessing, and mentorship of project 2000 diploma students. Nurse Education Today, 14(1) 44-51 (1994).

210. ENB&WNB, Teaching and Assessing in Clinical Practice, London, 1988.

211. Crotty, M. M., and Bignell, A. R., Using a quality assurance model to evaluate ENB course 998 'Teaching and Assessing in Clinical Practice'. Nurse Education Today, 8 332-340 (1988).

212. Spool, M. D., Training Programs for Observers of Behaviour: a Review. Personnel Psychology, 31 853-889 (1978).

213. Cline, D. W., and Garrard, J. N., Evaluation of the SAID teaching programme. American Journal of Psychiatry, (130) 582-585 (1973).

214. Mash, E. J., and Makohoniuk, G., The effects of prior information and behavioural predictability on observer accuracy. Child Development, 46 513-519 (1975).

215. Newble, D. I., Hoare, J., and Sheldrake, P. F., The selection and training of examiners for clinical examinations. Medical Education, 14 345-349 (1980).

216. Latham, G. P., Wexley, K. N., and Pursell, E. D., Training managers to minimise rating errors in the observation of behaviour. Journal of Applied Psychology, 1975(60) 550-555 (1975).

217. Cronin-Stubbs, D., and Mathews, J. J., A Clinical Performance Evaluation Tool for a Process-Oriented Nursing Curriculum. Nurse Educator, 7(4) 24-29 (1982).

218. Loustau, A., et al., Evaluating Student's Clinical Performance: Using Videotape to Establish Rater Reliability. Journal of Nursing Education, 19(7) 10-17 (1980).

219. Wexley, K. N., Sanders, R. E., and Yukl, G. A., Training interviewers to eliminate contrast effects in employment interviews. Journal of Applied Psychology, 50 233-236 (1973).

220. Goldstein, A. P., and Sorcher, M., Changing Supervisor Behaviour. Pergammon Press, New York, 1974.

221. Watson, N. A., Mentoring today - the students' views. An investigative case study of pre-registration nursing students' experiences and perceptions of mentoring in one theory/practice module of the CFP on a project 2000 course. Journal of Advanced Nursing, 29(1) 254-262 (1999).

222. Gray, M., and Smith, L., The professional socialisation of diploma of higher education in nursing students (Project 2000): a longitudinal study. Journal of Advanced Nursing, 29(3) 639-647 (1999).

223. Donovan, J., The concept and role of mentor. Nurse Education Today, 10 294-298 (1990).

224. Walsh, C. R., and Clements, C. A., Attributes of mentors as perceived by orthopaedic nurses. Orthopaedic Nursing, 14(3) 49-56 (1995).

225. Darling, L. A., What do Nurses want in a mentor? Journal of Nursing Administration, (October) 42- 44 (1984).

226. Earnshaw, G. J., Mentorship: the students' views. Nurse Education Today, 15 274-279 (1995).

227. Morton-Cooper, A., and Palmer, A., Mentoring, Preceptorship and Clinical Supervision. Blackwell Science, Oxford, 2000.

228. Morle, K. M. F., Mentorship - is it a case of the emperors new clothes or a rose by any other name? Nurse Education Today, 10 66-69 (1990).

229. Angelini, D. J., Mentoring in the Career Development of Hospital Staff Nurses: Models and Strategies. Journal of Professional Nursing, 11(2) 89-97 (1995).

230. Burnard, P., The student experience: adult learning and mentorship revisited. Nurse Education Today, 10 349-354 (1990).

231. Atkins, S., and Williams, A., Registered nurses' experience of mentoring undergraduate nursing students. Journal of Advanced Nursing, 21 1006 - 1015 (1995).

232. Andrews, M., and Wallis, M., Mentorship in nursing: a literature review. Journal of Advanced Nursing, 29(1) 201-207 (1999).

233. Jowett, S., Walton, I., and Payne, S., The Introduction of Project 2000: Early perspectives from the students. NFER, London, 1992.

234. Wright, C. M., An innovation in a diploma program: the future potential of mentorship in nursing. Nurse Education Today, 10 355-359 (1990).

235. Rogers, P., and Lawton, C., Self-assessment of project 2000 supervision. Nursing Times, 91(39) 42-45 (1995).

236. Darling, L. A., Mentors and mentoring. Journal of Nursing Administration, (March) 42-43 (1985).

237. Prestholdt, C. O., Modern Mentoring: strategies for developing contemporary nursing leadership. Nursing Administration Quarterly, 15(1) 20-27 (1990).

238. Fields, W. L., Mentoring in nursing: a historical approach. Nursing Outlook, 39(6) 257-261 (1991).

239. Armitage, P., and Burnard, P., Mentors or Preceptors? Narrowing the theory-practice gap. Nurse Education Today, 11 225-229 (1991).

240. Hagerty, B., A second look at mentors. Nursing Outlook, 34(1) 16-24 (1986).

241. Fowler, J., The organisation of clinical supervision within the nursing profession: a review of the literature. Journal of Advanced Nursing, 23 471-478 (1996).

242. Maggs, C., Mentorship in nursing and midwifery education: issues for research. Nurse Education Today, 14(1) 22-29 (1994).

243. Barlow, S., Impossible Dream. Nursing Times, 87(1) 53-54 (1991).

244. Infante, M. S., Forbes, E. J., and Houldin, A. D., A clinical teaching project: examination of a clinical teaching model. Journal of Professional Nursing, 5 132-139 (1989).

245. Kirkpatrick, H., et al., A collaborative model for the clinical education of baccalaureate students. Journal of Advanced Nursing, 16 101-107 (1991).

246. Butterworth, T., and Faugier, J., Clinical Supervision and Mentorship in Nursing. Chapman & Hall, London, 1992.

247. Cameron-Jones, M., Looking for Quality and Competence in Teaching. *In* Professional Competence and Quality Assurance in the Caring Professions, vol. 1, edited by R. Ellis. Croom Helm, London, 1988, pp. 58-83.

248. Cameron-Jones, M., The role of the Nurse Mentor. Proceedings of the Mentorship in Nurse and Midwife Education. Edinburgh, 1996.

249. Butterworth, T., Clinical Supervision as an emerging idea in nursing. *In* Clinical Supervision and Mentorship in Nursing, edited by T. Butterworth and J. Faugier. Chapman and Hall, London, 1992, pp. 3-17.

250. ENB, Preparation of Teachers, Practitioners/Teachers, Mentors and Supervisors in the Context of Project 2000. English National Board for Nursing, Midwifery, and Health Visiting., London, 1989.

251. UKCC, Framework of standards for the preparation of lecturers of nursing, midwifery and health visiting. UKCC, London, 1997.

252. Davis, J. A., You need someone older and wiser. American Journal of Nursing, (Oct.) 1290-1291 (1984).

253. Editorial, Consider This: What do nurses want in a mentor? Journal of Nursing Administration, 26(4) 6,7 (1996).

254. Cameron-Jones, M., and O'Hara, P., Three decisions about nurse mentoring. Journal of Nursing Management, 4 45-50 (1996).

255. Daloz, L., Effective Teaching and Mentoring. Jossey Bass, San Francisco, 1986.

256. Severinsson, E. I., The concept of supervision in psychiatric care - compared with mentorship and leadership. A review of the literature. Journal of Nursing Management, 2 271-278 (1994).

257. Spouse, J., The effective mentor: a model for student centred learning in clinical practice. Nursing Times Research, 1(2) 120-133 (1996).

258. Anforth, P., Mentors, not assessors. Nurse Education Today, 12 299-302 (1992).

259. Morris, N., John, G., and Keen, T., Learning the ropes. Nursing Times, 84(46) 24-27 (1988).

260. Bodley, J., Adapting supervision strategies to meet the challenges of future mental health nursing practice. Nurse Education Today, 11 378-386 (1991).

261. May, K. M., Meleis, A. I., and Winstead-Fry, P., Mentorship for Scholarliness: opportunities and dilemmas. Nursing Outlook, Jan 22-28 (1982).

262. Hunt, D. M., and Michael, C., Mentorship: a career training and development tool. Academy of Management Review, 8(3) 475-485 (1983).

263. Cahill, H. A., A qualitative analysis of student nurses' experiences of mentorship. Journal of Advanced Nursing, 24(4) 791-9 (1996).

264. Bidwell, A. S., and Brasler, M. L., Role modelling versus mentoring in nursing education. Image, 21(1) 23-25 (1989).

265. Spouse, J., The effective mentor: a model for student centred learning. Nursing Times, 92(13) 32-35 (1996).

266. Pelletier, D., and Duffield, C., Is there enough mentoring in nursing? The Australian Journal of Advanced Nursing, 11(4) 6-11 (1994).

267. Atkinson, R. L., et al., Introduction to Psychology. Harcourt, Brace Jovanovich, Orlando, Florida, 1990.

268. Dawson, K. P., Attitude and assessment in nurse education. Journal of Advanced Nursing, 17(4) 473-9 (1992).

269. Baillie, L., Factors affecting student nurses' learning in community placements: a phenomenological study. Journal of Advanced Nursing, 18 1043-1053 (1993).

270. Calderone, A. B., Computer assisted instruction: Learning, attitude and modes of instruction. Computers in Nursing, 12(3) 164-170 (1994).

271. Harvey, T. J., and Vaughan, J., Student nurse attitudes towards different teaching/learning methods. Nurse Education Today, 10 181-185 (1990).

272. Oppenheim, A. N., Questionnaire Design, Interviewing, and Attitude Measurement. Pinter Publishers, London, 1992.

273. Gray, C., Patients' perception of anxiety. Surgical Nurse, 2(3) 12-17 (1989).

274. Martin, D., Pre-operative visits to reduce patient anxiety: a study. Nursing Standard, 10(23) 33-38 (1996).

275. Bell, M. L., Learning a complex nursing skill: student anxiety and the effect of preclinical skill evaluation. Journal of Nursing Education, 30(5) 222-226 (1991).

276. Lloyd Jones, M., and Akehurst, R., Clinical placements for student nurses and midwives: an evaluation of costs and practices. Scharr, University of Sheffield, Sheffield, 1997.

277. Yonge, O., Krahn, H., and Trojan, L., Perceptions of preceptors toward preceptorship in nursing undergraduate education programs. AARN, 50(9) 14-16 (1994).

278. Hayes, E., Helping preceptors mentor the next generation of nurse practitioners. Nurse Practitioner, 19(6) 62-66 (1994).

279. Wilson, A., A new role for staff nurses: the effects of supernumerary status. Professional Nurse, 5(2) 105-106 (1989).

280. Twinn, S., Issues in the supervision of health visiting practice: an agenda for debate. *In* Clinical Supervision and Mentorship in Nursing, edited by T. Butterworth and J. Faugier. Chapman and Hall, London, 1992.

281. Craddock, E., Developing the facilitator role in the clinical area. Nurse Education Today, 13 217-224 (1993).

282. D'A Slevin, O., and Lavery, M., Self-directed learning and student supervision. Nurse Education Today, 15 274-279 (1995).

283. Price, B., Student nurse appraisal. Nursing Times, 85(40) 42-44 (1989).

284. Bradley, J., Is clinical progress assessment fair? Senior Nurse, 7(5) 14-16 (1987).

285. Neary, M., Defining the role of assessors, mentors and supervisors: part 1. Nursing Standard, 11(42) 34-39 (1997).

286. Hallett, C. E., et al., The implementation of project 2000 in the community: a new perspective on the community nurse's role. Journal of Advanced Nursing, 21 1159-1166 (1995).

287. Hamill, C., The phenomenon of stress as perceived by Project 2000 student nurses: a case study. Journal of Advanced Nursing, 21(3) (1995).

288. Naish, J., Scarce Resources. Nursing Times, 9(2) 32-33 (1996).

289. Smith, P., Nurses' emotional labour. Nursing Times, 85(47) 49-51 (1989).

290. Lindop, E., A comparative study of stress between pre- and post-Project 2000 students. Journal of Advanced Nursing, 29(4) 967-973 (1999).

291. Kleehammer, K., and Keck, J. F., Nursing students' perceptions of anxiety-producing situations in the clinical setting. Journal of Nursing Education, 29(4) 183-187 (1990).

292. Wood, V., An examination of selected current student nurse issues. Nurse Education Today, 12 403-408 (1992).

293. Goldsmith, J. W., Effect of learner variables, media attributes, and practice conditions on psychomotor task performance. Western Journal of Nursing Research, 6(2) 229-240 (1984).

294. Eysenck, M., and Keane, M., Cognitive Psychology. Lawrence Erlbaum Associates, Hove, 1995.

295. Elliott, D., Comparison of three instruments for measuring patient anxiety in a coronary care unit. Intensive and Critical Care Nursing, 9 195-200 (1993).

296. Farina, F., et al., Predictors of anxiety towards computers. Computers in Human Behaviour, 7 263-267 (1991).

297. McInerney, and Valentina, Definitely not just another computer anxiety instrument: the development and validation of CALM: computer anxiety and learning measure. Proceedings of the Annual Stress and Anxiety Research Conference. Madrid, 1994.

298. Spielberger, C. D., et al., Manual for the State-Trait Anxiety Inventory. Mind Garden, Palo Alto California, 1983.

299. Streiner, D. L., A Checklist for Evaluating the Usefulness of Rating Scales. Canadian Journal of Psychiatry, 38 140-148 (1993).

300. Mindgarden.com, http://www.Mindgarden.com, 2001.

301. Jinks, G. H., Making the most of practice placements: what the nurse teacher can do to maximise the benefits for students. Nurse Education Today, 11 127-133 (1991).

302. Davis, C., Davis, B. D., and Burnard, P., Use of the QSR.NUD.IST computer program to identify how clinical midwife mentors view their work. Journal of Advanced Nursing, 26 833-839 (1997).

303. Neary, M., Defining the role of assessors, mentors and supervisors: part 2. Nursing Standard, 11(43) 34-38 (1997).

304. Jackson, F., Problems in assessing nursing students. Nursing Times, 79(23) 33-34 (1983).

305. Quinn, F. M., The Principles and Practice of Nurse Education. Chapman & Hall, Glasgow, 1995.

306. Bandura, A., Social Foundations of Thought and Action: a Social Cognitive Theory. Prentice-Hall, Englewood Cliffs, New Jersey, 1986.

307. Murdock, J. E., and Neafsey, P. J., Self-efficacy measurements: an approach for predicting practice outcomes in continuing education? Journal of Continuing Education in Nursing, 26(4) 158-165 (1995).

308. Pajares, F., Self-efficacy in academic settings. Proceedings of the Annual Meeting of the American Educational Research Association. San Francisco, 1995.

309. Hodge, M. B., Do anxiety, math self-efficacy, and gender affect nursing students' drug dosage calculations? Nurse Researcher, 24(4) 36-41 (1999).

310. Haycock, L. A., McCarthy, P., and Skay, C. L., Procrastination in college students: the role of self-efficacy and anxiety. Journal of Counselling and Development, 76(3) 17-24 (1998).

311. Zhang, Y., and Espinoza, S., Affiliations of computer self-efficacy and attitudes with need for learning computer skills. Journal of Educational Computing Research, 17(4) 371-383 (1997).

312. Gorrell, J., Some contributions of self-efficacy research to self-concept theory. Journal of Research and Development in Education, 23(2) 73-81 (1990).

313. Eastman, C., and Marzillier, J. S., Theoretical and methodological difficulties in Bandura's self-efficacy theory. Cognitive Ther Res, 8 213-229 (1984).

314. Maddux, J. E., and Stanley, M. A. e., Self-efficacy theory in contemporary psychology (special issue). Journal of Social and Clinical Psychology, 4(3) (1986).

315. Brady, T. J., Do common arthritis self-efficacy measures really measure self-efficacy? Arthritis Care and Research, 10(1) 1-8 (1997).

316. Lorig, K., et al., Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis and Rheumatology, 32 37-44 (1989).

317. Mager, R. F., No self-efficacy, no performance. Training, 1992(April) 32-36 (1992).

318. Davis, P., et al., Evaluation of a rheumatoid arthritis patient education program: impact on knowledge and self-efficacy. Patient Education and Counselling, 24(1) 55-61 (1994).

319. Goldenberg, D., Iwasiw, C., and MacMaster, E., Self-efficacy of senior baccalaureate nursing students and preceptors. Nurse Education Today, 17 303-310 (1997).

320. Harvey, V., and McMurray, N., Self-efficacy: a means of identifying problems in nursing education and careers progress. International Journal of Nursing Studies, 31(5) 471-485 (1994).

321. Vispoel, W., and Chen, P., Measuring self-efficacy: the state of the art. Proceedings of the Annual Meeting of the American Educational Research Association. Boston, 1990.

322. O'Halloran, V. E., et al., Improving self-efficacy in nursing research. Clinical Nurse Specialist, 10(2) 83-87 (1996).

323. Schlessel, J. S., et al., CPR knowledge, self-efficacy, and anticipated anxiety as functions of infant/child CPR training. Annals of Emergency Medicine, 25(5) 618-623 (1995).

324. Draper, S. W., et al., Integrative evaluation: an emerging role for classroom studies of cal. Computers and Education, 26 17-32 (1996).

325. Myrick, F., and Awrey, J., The effect of preceptorship on the clinical competency of baccalaureate student nurses: a pilot study. Canadian Journal of Nursing Research, 20(3) 29-43 (1988).

326. Schwartz, C. E., et al.,Measuring self-efficacy in people with multiple sclerosis: a validation study. Archives of Physical Medicine and Rehabilitation, 77(4) 394-8 (1996).

327. Altmaier, E. M., et al.,Role of self-efficacy in rehabilitation outcome among chronic low back pain patients. Journal of Counselling in Psychology, 40 335-339 (1993).

328. Booth, K., Providing support and reducing stress: a review of the literature. *In* Clinical Supervision and Mentorship in Nursing, edited by T. Butterworth and J. Faugier. Chapman and Hall, London, 1992, pp. 50-59.

329. Clifford, C.,Nurse teacher's clinical work: a survey report. Journal of Advanced Nursing, 23 603-611 (1996).

330. Fish, D.,Quality Mentoring for Student Teachers: a principled approach to practice. D. Fullan Publishers, , London, 1995.

331. Cerinus, M., and Ferguson, C.,Preparing nurses for preceptorship. Nursing Standard, 8(36) 34-8 (1994).

332. Westra, R., and Graziano, M.,Preceptors: a comparison of their perceived needs before and after the preceptor experience. Journal of Continuing Education in Nursing, 23 212-215 (1992).

333. Brownrigg, M.,Spare a thought for the mentors. Nursing Standard, 6(48) 52-53 (1992).

334. Laurent, C.,On Hand To Help. Nursing Times, 84(46) 29-30 (1988).

335. Williams, P. L., and Webb, C.,Clinical supervision skills: a Delphi and critical incident technique study. Medical Teacher, 16(2/3) 139-157 (1994).

336. James, M., and Turner, P.,Manymedia or is it multimedia? Information Technology in Nursing, 5(1) 8-9 (1993).

337. King, T.,Authoring for distance learning. The CTISS File, (July) 30-32 (1994).

338. Lekan, H., Index to Computer Assisted Instruction. Sterling Institute, Boston, 1970.

339. Lewis, M. J., et al.,A review of evaluative studies of computer-based learning in nursing education. Nurse Education Today, 21 26-37 (2001).

340. Rowntree, D.,Teaching through self-instruction - a Practical Handbook for Course developers. Kogan Page, , London, 1986.

341. Ehrman, S. C., The Flashlight Project: Spotting an Elephant in the Dark, vol. 1998, 1997.

342. Atkins, M. J.,Theories of learning and multimedia applications: an overview. Research Papers in Education, 8(2) 251-271 (1993).

343. Lowe, D., and Hall, W.,Hypermedia and the Web: an engineering approach. Wiley, , London, 1999.

344. Laurillard, D.,Re-thinking University Teaching. Routledge, , London, 1993.

345. Ribbons, R.,The use of computers as cognitive tools to facilitate higher order thinking skills in nurse education. Computers in Nursing, 16(4) 223-228 (1998).

346. Conole, G., and Oliver, M.,A pedagogical framework for embedding C&IT into the curriculum. Association for Learning Technology Journal, 6(2) 4-16 (1998).

347. Reeve, J.,InimiCAL or PedagogiCAL. The British Journal of Healthcare Computing & Information Management, 11(9) 22-23 (1994).

348. Procter, P. M.,Nurses, Computers, and Information Technology. Chapman Hall, , London, 1992.

349. Gunn, C.,Cal evaluation: what questions are being answered? A response to the article "integrative evaluation" by Draper et al. Computers and Education, 27 157-160 (1996).

350. Draper, S. W.,Prospects for summative evaluation of cal in higher education. Association for Learning Technology Journal, 5(1) 33-39 (1997).

351. James, M., and Turner, P.,Computer assisted learning - any progress? Information Technology in Nursing, 4(4) 10-11 (1992).

352. Clark, R. E., and Sugrue, B. M.,North American disputes about research on learning from media. International Journal of Educational Research, 14(6) 507-520 (1990).

353. Segal, J.,An evaluation of a teaching package constructed using a web-based lecture recorder. Association for Learning Technology Journal, 5(3) 32-42 (1997).

354. Jesshope, C., Shafarenko, A., and Slusanchi, H., Low bandwidth multimedia tools for Web-based lecture publishing., vol. 1998. The Computer Forum, 1998.

355. Gee, P. R., et al.,Development and evaluation of a computer assisted instruction package in clinical pharmacology for nursing students. Computers in Nursing, 16(1) 37-44 (1998).

356. Johnson, M. I.,Computer-based interactive tutorial versus traditional lecture for teaching introductory aspects of pain. Association for Learning Technology Journal, 5(3) 22-31 (1997).

357. Kenny, J., Too much skimmed, too little milked. *In* Times Education Supplement, London, 1996, pp. 8.

358. Somekh, B.,The evaluation of teaching with computers. The CTISS File, (10) 32-39 (1990).

359. Jayasuriya, R., and Caputi, P., Computer attitude and computer anxiety in nursing: validation of an instrument using an Australian sample. Computers in Nursing, 14(6) 340-345 (1996).

360. Selwyn, N.,Students' attitudes toward computers: validation of a computer attitude scale for 16-19 education. Computers and Education, 28(1) 34-41 (1997).

361. Kay, R.,An exploration of theoretical and practical foundations for assessing attitudes toward computers: the Computer Attitude Measure. Computers in Human Behaviour, 9 371-386 (1993).

362. Woodrow, J. J.,A comparison of four computer attitude scales. Journal of Educational Computing Research, 7(2) 165-187 (1991).

363. Schwirian, P. M., et al.,Computers in nursing practice: a comparison of the attitudes of nurses and nursing students. Computers in Nursing, 7(4) 168-177 (1989).

364. Schare, B. L., et al.,The effects of interactive video on cognitive achievement and attitude toward learning. Journal of Nursing Education, 30(3) 109-113 (1991).

365. Hamby, C. S.,A study of the effects of computer assisted instruction on the attitude and achievement of vocational nursing students. Computers in Nursing, 4(3) 109-113 (1986).

366. Lowdermilk, D. L., and Fishel, A. H.,Computer simulations as a measure of nursing students' decision making skills. Journal of Nursing Education, 30(1) 34-39 (1991).

367. Jensh, R.,Use of interactive video programs in education in basic medical sciences. Journal of Medical Education, 62 7-14 (1987).

368. Soled, S. W., et al., The effects of interactive video on cognitive achievement and attitude toward learning. Proceedings of the Annual Meeting of the American Educational Research Association, 1-6. San Francisco, 1989.

369. Saranto, K., Leino-Kilpi, H., and Isoaho, H.,Learning environment in information technology: the views of student nurses. Computers in Nursing, 15(6) 324-332 (1997).

370. VanDover, L., and Boblin, S.,Student nurse computer experience and preferences for learning. Computers in Nursing, 9(2) 75-79 (1991).

371. Day, R., and Payne, L.,Computer managed instruction: an alternative teaching strategy. Journal of Nursing Education, 26(1) 30-36 (1987).

372. Thede, L. Q., Taft, S., and Coeling, H.,Computer assisted instruction: a learner's viewpoint. Journal of Nursing Education, 33(7) 299-305 (1994).

373. Baldwin, D., Johnson, J., and Hill, P.,Student satisfaction with classroom use of computer assisted instruction. Nursing Outlook, 42(4) 188-192 (1994).

374. Jones, B. T., and McCormac, K.,Empirical evidence shows that measuring users' opinions is not a satisfactory way of evaluating computer-assisted learning in nurse education. International Journal of Nursing Studies, 29(4) 411-425 (1992).

375. Brosnan, M. J., and Davidson, M. J.,Computerphobia - is it a particularly female phenomenon? The Psychologist, 7(2) 73-78 (1994).

376. Morrow, P. C., Prell, E. R., and McElroy, J. C.,Attitudinal and behavioural correlates of computer anxiety. Psychology Reports, 59 1199-1204 (1986).

377. Staggers, N., and Mills, M. E.,Nurse - Computer Interaction: Staff Performance Outcomes. Nursing Research, 43(3) 144-150 (1994).

378. Abla, S.,The who, what, where, when and how of computer education. Computers in Nursing, 3 114-117 (1995).

379. Billings, D. M., and Cobb, K. L.,Effects of learning style preferences, attitude and GPA on learner achievement using computer assisted interactive videodisc instruction. Journal of Computer Based Instruction, 19(1) 12-16 (1992).

380. Koch, E. W., Rankin, J. A., and Stewart, R.,Nursing students' preferences in the use of computer assisted learning. Journal of Nursing Education, 29(3) 122-126 (1990).

381. Ellis, D.,Multimedia in education and training: promises and challenges. The CTISS File, (16) 18-21 (1993).

382. Somekh, B.,Designing software to maximise learning. Association for Learning Technology Journal, 4(3) 4-16 (1996).

383. Miller, R. L.,Ten Good Reasons. Learning Benefits of Interactive Technologies. The Videodisc Monitor, (Feb 1990) 15-16 (1990).

384. Smeaton, A., and Keough, G.,An analysis of the use of virtual delivery of undergraduate lectures. Computers and Education, 32 83-94 (1999).

385. DeAmicis, P.,Interactive videodisc instruction is an alternative method for learning and performing a critical nursing skill. Computers in Nursing, 15(3) 155-158 (1997).

386. Lo, R., et al.,The development and evaluation of a computer-aided diabetes education program. Australian Journal of Advanced Nursing, 13(4) 19-27 (1996).

387. Blackley, J. A., and Curran-Smith, J.,Teaching community health nursing by distance methods: development, process and evaluation. Journal of Continuing Education for Nurses, 29(4) 148-153 (1998).

388. Russell, T. J., The No Significant Difference Phenomenon, vol. 2001, 1999.

389. Kulik, J. A., and Kulik, C. C.,Review of recent research literature on computer-based instruction. Contemporary Educational Psychology, 12 222-230 (1987).

390. Cohen, P. A., and Dacanay, L. S.,A meta-analysis of computer based instruction in nursing education. Computers in Nursing, 12(2) 89-97 (1994).

391. Belfry, M. J., and Winne, P. H.,A review of the effectiveness of computer assisted instruction in nursing education. Computers in Nursing, 6(2) 77-85 (1988).

392. Pedlar, J.,A preliminary evaluation of a national programme of computer assisted learning for oral surgery. Computers in Teaching Initiative Centre For Medicine Update, 4(1) 13 (1993).

393. Koch, B., and Rankin, J., Computers and their Applications in Nursing. Harper & Row, London, 1987.

394. Jones, R. V. H., et al.,CAL in cancer nursing: an adult learning approach. Information Technology in Nursing, 2(3) 34-35 (1990).

395. Umlauf, M. G.,How to provide around-the-clock CPR certification without losing any sleep. Journal of Continuing Education in Nursing, 21(6) 248-251 (1990).

396. Clark, R. E.,Confounding in educational computing research. Journal of Educational Computing Research, 1(2) 137-148 (1985).

397. Mayes, J. T.,Commentary - impact of cognitive theory on the practice of courseware authoring. Journal of Computer Assisted Learning, 9 222-228 (1993).

398. Britain, S., and Liber, O., A framework for the pedagogical evaluation of virtual learning environments, vol. 1999. JISC Technology Applications Programme, http://www.jtap.ac.uk/reports/htm/jtap-041.html. Accessed 15 November 1999.

399. Rushby, N.,Quality criteria for multimedia. Association for Learning Technology Journal, 5(2) 18-30 (1997).

400. Gunn, C., CAL evaluation: future directions. Association for Learning Technology Journal, 5(1) 40-47 (1997).

401. Johnson, P., Human Computer Interaction. McGraw & Hill, , London, 1992.

402. Booth, P., An Introduction to Human-Computer Interaction. Lawrence Erlbaum Associates, Publishers., , Hove, 1992.

403. Preece, J., et al., Human-Computer Interaction. Addison-Wesley Publishing Company, , Wokingham, 1994.

404. Hicks, C., The randomised controlled trial: a critique. Nurse Researcher, 6(1) 19-31 (1998).

405. Tamblin, L., and Sleep, J., Randomized Controlled Studies. Surgical Nurse, 7(3) 10-13 (1994).

406. Getliffe, K., Developing a protocol for a randomised controlled trial: factors to consider. Nurse Researcher, 6(1) 5-17 (1998).

407. Eaton, N. M., Interactive Video. Surgical Nurse, (1993).

408. Pillay, H., Boles, W., and Raj, L., Personalizing the design of computer-based instruction to enhance learning. Association for Learning Technology Journal, 6(2) 17-32 (1998).

409. Richards, S., Educational software design: applying models of learning. Association for Learning Technology Journal, 4(3) 17-20 (1996).

410. Jones, P., An enthusiast's view of CAL. Information Technology in Nursing, 6(2) 4-6 (1994).

411. Roberts, J. D., While, A. E., and Fitzpatrick, J. M., Simulation: current status in nurse education. Nurse Education Today, 12 409-415 (1992).

412. Pond, E. F., Bradshaw, M. J., and Turner, S. L., Teaching strategies for critical thinking. Nurse Educator, 16(6) 18-22 (1991).

413. Edward, N. S., Computer based laboratory simulation: evaluations of student perceptions. Association for Learning Technology Journal, 4(3) 41-53 (1996).

414. Henry, S. B., and Holzemer, W. L., The relationship between performance on computer-based clinical simulations and two written methods of evaluation: cognitive examination and self-evaluation of expertise. Computers in Nursing, 11(1) 29-34 (1993).

415. Neelamkavil, F., Computer Simulation and Modelling. John Wiley and sons Ltd., , Chichester, 1987.

416. Hmelo, C. E., Computer assisted instruction in health professions education: a review of the published literature. Journal of Educational Technology Systems, 18(2) 83-101 (1989).

417. Miller, M. D., Simulations in Medical Education. Medical Teacher, 9 35-41 (1987).

418. Tam, M., Constructivism, Instructional Design, and Technology: implications for transforming distance learning. Educational Technology and Society, 3(2) 1436-4522 (2000).

419. Wright, D. J.,The use of multimedia computers and software in nurse education. Health Informatics, 1 101-107 (1995).

420. Arnold, S., et al., Constructing and Implementing Multimedia Teaching Packages. University of Glasgow, Glasgow, 1994.

421. Lebow, D., Constructivist values for systems design: five principles toward a new mindset. Educational Technology Research and Development, 41 4-16 (1993).

422. Rheaume, J., L'Enseignement des hypermedias pedagogiques. Lille, France, 1993.

423. King, A., and Honeybone, A., Needs before means: the dialectics of learning and technology. Association for Learning Technology Journal, 4(2) 4-16 (1996).

424. Collis, B.,Tele-learning in a Digital World: The future of Distance Learning. International Thomson Computer Press, , London, 1996.

425. McAteer, E., and Shaw, R., The Design of Multimedia Learning Programs. University of Glasgow, Glasgow, 1995.

426. Laurillard, D., The Changing University, vol. 2001, 1999.

427. Foltz, P. W., Comprehension, Coherence and Strategies in Hypertext and Linear text. *In* Hypertext and cognition, edited by J.-F. Rouet, et al. Lawrence Erlbaum Associates, Hillsdale, N.J, 1996.

428. Schwier, R. A., Issues in emerging interactive technologies. *In* Instructional technology: Past, present and future, edited by G. J. Anglin. Libraries Unlimited inc., Englewood, C.O., 1995, pp. 119-127.

429. Shuell, T. J.,The role of the student in learning from instruction. Contemporary Educational Psychology, 13(3) 276-295 (1988).

430. Ertmer, P. A., and Newby, T. J., Behaviourism, Cognitivism, Constructivism: comparing critical features from an instructional design perspective. Performance Improvement Quarterly, 6(4) 50-70 (1993).

431. Chandra, P., Hinton, T., and Phillips, N., Design Considerations for the development of interactive video in nurse education. Computers and Education, 15(1) 183-188 (1990).

432. Gagne, R. M.,The Conditions of Learning. Holt, Rinehart and Winston., , London, 1967.

433. Ohlsson, S.,Impact of cognitive theory on the practice of courseware authoring. Journal of Computer Assisted Learning, 9 194-221 (1993).

434. Lyons, J., and Milton, J.,Recognizing through feeling: a physical and computer simulation based on educational theory. Computers in Nursing, 17(3) 114-119 (1999).

435. MacDonald, B., The educational potential of computer-assisted learning: qualitative evidence about student learning'. UNCAL, 1976.

436. Jonassen, D. H.,Objectivism versus constructivism: do we need a new philosophical paradigm? Educational Technology Research and Development, 39(3) 5-14 (1991).

437. Reigeluth, C. M.,A new paradigm of ISD? Educational Technology, (May-June) 13-20 (1996).

438. James, M., and Turner, P.,Multimedia - doing it, not just talking about it. Information Technology in Nursing, 5(4) 5-7 (1993).

439. Ribbons, R. M., Guidelines for developing interactive multimedia: applications in nurse education. Computers in Nursing, 16(2) 109-114 (1998).

440. Carrigan, T., Tooled-up fools. MacUser, 11(18) 71 (1995).

441. Jones, B., Masters of CBL. *In* Multimedia and Knowledge Based Systems Newsletter, 1995, pp. 2.

442. McBeath, C., and Atkinson, R.,Curriculum, instructional design and the technologies: Planning for educational delivery. Australian Journal of Educational Technology, 8(2) http://cleo.murdoch.edu.au/gen/aset/ajet/ajet8/su92p119.html, 119-131 (1992).

443. Lyte, V. J., and Kerr, J., Developing interactive multi-media for health care education and training. Health Informatics, 2 139-145 (1996).

444. Hill, M. N., Staffing a Distance Learning Team: whom do you really need? Online Journal of Distance Learning Administration, 1(1, Spring 1998) http://www.westga.edu/~distance/hill.html, (1998).

445. McAteer, E., and Shaw, R., Courseware in Higher Education Evaluation 1: Planning, Developing and Testing. The EMASHE group, University of Glasgow, Glasgow, 1994.

446. Dickens, J., and Sherwood-Edwards, M., Multimedia and Copyright: some obstacles on the yellow brick road. Proceedings of the BCS annual Conference 'Multimedia Systems and Applications'. Leeds, 1993.

447. Bassett, C.,Rights Issues. MacUser, 10(23) 87-88 (1994).

448. Burke, L. M., Effective evaluation of CAL software packages. Information Technology in Nursing, 5(2) 5-7 (1993).

449. Essex, M ,Evaluation of computer assisted learning packages. Health Informatics, 2 78-80 (1996).

450. Jones, B. T., and McCormac, K.,The structure and process of evaluating computer assisted learning in nurse education (N-CAL): human-computer interaction. Information Technology in Nursing, 3(1) 8-14 (1991).

451. Leonard, J.,Interacting: multimedia and health. Health Education Authority, Exeter, 1994.

452. Phillips, R.,The developers handbook to interactive multimedia. Kogan Page, London, 1997.

453. Tessmer, M.,Planning and Conducting Formative Evaluations. Kogan Page, London, 1993.

454. Chanier, T.,Evaluation in a project life-cycle: the hypermedia CAMILLE project. Association for Learning Technology Journal, 4(3) 54-68 (1996).

455. Thiagarajan, S.,Instructional product verification: 20 questions and 200 speculations. Educational Communications and Technology Journal, 26(2) 133-142 (1978).

456. Jones, J. A.,The Verbal Protocol: a research technique for nursing. Journal of Advanced Nursing, 14 1062-1070 (1989).

457. Green, A.,Verbal protocol analysis. The Psychologist, 8(3) 126-129 (1995).

458. Zelen, M.,A new design for randomized controlled trials. The New England Journal of Medicine, 300 1242-1245 (1979).

459. Jadad,A.R., Randomised Controlled Trials, vol. 2002. http://www.bmjpg.co.uk/rct/index.html BMJ, 1998.

460. Roland, M., and Torgerson, D. J.,What are pragmatic trials? British Medical Journal, 316 285 (1998).

461. Black, S.,Editorial: Randomised Controlled Trials. Nurse Researcher, 6(1) 3-4 (1998).

462. Harris, H.,Semi-structured interview schedules for gathering psychosocial health data. Nurse Researcher, 5(1) 73-85 (1997).

463. Chapple, A.,The use of telephone interviewing for qualitative research. Nurse Researcher, 6(3) 85-93 (1999).

464. Cohen, L., and Manion, L.,Research Methods in Education. Routledge, London, 1994.

465. Armstrong, M. L., and DeWit, S. C.,Choosing software: the importance of instructional design. Nurse Educator, 10(4) 13-17 (1985).

466. Heacock, P., Souder, E., and Chastain, J.,Subjects, Data, and Videotapes. Nursing Research, 45(6) 336-338 (1996).

467. Barab, S., and Duffy, T. M., From Practice Fields to Communities of Practice. *In* Theoretical Foundations of Learning Environments, edited by D. Jonassen and S. Land. Erbaum, New Jersy, 1999.

468. Docherty, C., Developing mentors' skills in assessing student nurses - a multimedia approach. Proceedings of the Mentorship in Nurse and Midwife Education, 22-28. Edinburgh, 1996.

469. Docherty, C., The design and evaluation of a multimedia program to improve mentors' assessments of student nurses. Proceedings of the Nurse Education Tomorrow. Durham, 1998.

470. Cohen, J.,Statistical Power Analysis for the Behavioural Sciences. Academic Press, New York, 1977.

471. Jaccard, J., and Becker, J.,Statistics for the Behavioural Sciences. Wadsworth, Belmont CA, 1990.

472. Hillier, V., and Gibbs, A.,Calculations to determine sample size. Nurse Researcher, 3(4) 27-34 (1996).

473. Blacktop, J.,A discussion of different types of sampling techniques. Nurse Researcher, 3(4) 5-15 (1996).

474. Pickering, J.,Teaching on the internet is learning. Active Learning, 2 9-12 (1995).