





Project on the Effectiveness of Problem Based Learning (PEPBL)

RESEARCH REPORT

July 2004

Problem Based Learning: An exploration of the method and evaluation of its effectiveness in a continuing nursing education programme.

Mark Newman

This study was completed as part of the Project on the Effectiveness of Problem Based Learning (PEPBL) funded by the Economic & Social Research Council's Teaching & Learning Research Programme.

Mark Newman was the grant holder and principal investigator for the Project on the Effectiveness of Problem Based Learning. The grant was awarded and the project carried out whilst Mark Newman was employed as senior research fellow in the School of Lifelong Learning and Education & School of Health and Social Sciences at Middlesex University between January 2000 and July 2003.

Mark Newman's contact details:
Evidence for Policy and Practice Information and Coordinating Centre (EPPI- Centre)
Social Science Research Unit (SSRU)
Institute of Education
University of London
18 Woburn Square
London WC1H 0NR

Tel: +44 (0) 207 612 6575
Fax: +44 (0) 207 612 6400
E:mail – m.newman@ioe.ac.uk

The following people also made substantial contributions to this project

Name	Dept/ Institution	Contribution
Kate Ambrose	School of Health & Social Sciences, Middlesex University	Design of PBL programme, teacher on programme
Trevor Corner	School of Lifelong Learning and Education, Middlesex University	Institutional sponsorship, project support
Jeff Evans	Middlesex University Business School	Advice on study design, data collection and analysis
John Koushappas	School of Health & Social Sciences, Middlesex University	Development and management of project website, formatting project outputs for publication.
Phyl Morris Vincent	School of Health & Social Sciences, Middlesex University & Royal Free Hospital	Design of PBL programme, teacher on programme
Sheila Quinn	School of Health & Social Sciences, Middlesex University	Design of PBL programme, teacher on programme
Lesley Vernon	School of Health & Social Sciences, Middlesex University	Design of PBL programme, teacher on programme
Sarah Wallis	School of Health & Social Care, Oxford Brookes University*	Design of PBL programme, teacher on programme

*Formerly of School of Health & Social Sciences, Middlesex University

Published By:
School of Lifelong Learning & Education,
Middlesex University,
Trent Park, Bramley Rd
London. N14 4YZ

© Copyright
Mark Newman holds the copyright for the text of this report. The School of Life Long Learning & Education and the author give permission for users of the site to display and print the contents of the site for their own non-commercial use, providing that the materials are not modified, copyright and other proprietary notices contained in the materials are retained, and the source of the material is cited clearly following the citation details provided. Otherwise users are not permitted to duplicate, reproduce, re-publish, distribute, or store material from this website without express written permission.

ISBN: 1 85924 241 3

Report Contents

(A detailed list of the contents of is provided at the beginning of each part of the report)

	Page
Abstract	**6**
Part I Introduction, study methods	**8**
Part I contents page	9
Introduction	11
Study design and methods	23
Part II: Study contexts, Participants, curriculum theory and practice	**67**
Part II contents page	68
Introduction	70
The Specific institutional and disciplinary contexts of the study	76
Teacher and student cultures and relationships in the study programmes	80
Programme content and assessment in the study	102
The control (SGL) curriculum in operation	112
Design and organisation of the experimental (PBL) curriculum	115
Exploring differences between the two curricula -	139
Part III: Results discussion and conclusions	**147**
Part III contents page	148
Progress through the experiment sample size and power at the completion of the programme	151
Results: Learner satisfaction	154
Results - Skills Personal and Propositional Knowledge	166
Results self –directed learning -	183
Results teacher workload	190
Discussion	192
Conclusions	206
References	**225**

Acknowledgements

This study would not have been possible without the cooperation, support, and advice given by lots of different people. A particular debt is due to the teachers who took part in the study, Kate Ambrose, Lesley Vernon, Phyl Morris-Vincent, Sheila Quinn, and Sarah Wallis. Their agreement to expose their beliefs and practices as teachers to external public scrutiny exemplifies the highest standards of professionalism. That they were interested in and committed to improving their practice even in some cases after many years of practice as teachers provide us with excellent role models for life long learning. Thanks are also due to the students who were also subjects of the study and who did not complain too much about filling in so many questionnaires and being scrutinised in their classrooms.

The design and conduct of the Problem Based Learning model used in the study benefited from the advice of a number of people in particularly Danke Cleverley and Hilary Bebb (Anglia Polytechnic University), Liz Rideout (McMaster University), Sue Grandis (University of Southampton) Garth Long (Nursing & Midwifery Council), Kay Wilkie (University of Dundee), Alan Ryan (Whittington Hospital NHS Trust), Maureen Pennington-Lowe (Barnet & Chase Farm Hospitals NHS Trust). I would also like to thank the staff and students in the Nursing programmes at McMaster and Anglia Polytechnic Universities who allowed us to come and observe their Problem Based Learning Tutorials.

The design and conduct of the overall project and of the research itself benefited from the advice and input from a number of people in particular the project advisory group. Hamza Aumeer (Formerly the ENB), Carol Baxter (Middlesex University), Trevor Corner (Middlesex University)(Chair), Phil Davies (University of Oxford and The Cabinet Office), Jeff Evans (Middlesex University), John Kanefsky (ESRC Teaching & Learning Research Programme), John Stephenson (Middlesex University), Rena Papadopoulos (Middlesex University), Andrew Pollard (ESRC Teaching & Learning Research Programme & University of Cambridge), Janice Sigsworth (North Central London NHS Workforce Confederation), Peter Tymms (University of Durham) John Webb (Middlesex University) and the teachers named above.

A specific acknowledgement is due to Charles Engel one of the early pioneers of Problem Based Learning whose critical feedback on all aspects of the study design and indeed this report have no doubt resulted in a better study and product than would otherwise have been the case. Thanks also to Velda McCune (University of Edinburgh) for her comments on part II of the report.

Numerous people contributed to the conduct of the study. Thanks to Heather Cahill (University of York), Carolyn Gibbon (University of Lancaster), Fiona Heskin (Homerton College Cambridge), for acting as external independent markers. Thanks to the Internal Learning Resources Services team in the School of Health and Social Sciences at Middlesex University for their assistance in making the videos and David Atkins in particular for editing. Thanks to John Newman and Ann Pickering for assessing the videos of group performance. Thanks to Stephen Cook for providing supervision for the teachers. Thanks to Noel Entwistle and colleagues for their permission to use ASSIST and to Murray Fisher and colleagues for permission to use the Self Directed Learning Readiness Scale.

Thanks to John Kousahappas the ICT manager in the School of Health & Social Sciences at Middlesex University for his help in developing and maintaining the projects webpages

The ESRC Teaching & Learning Research Programme funded the research upon which this report is based. I would like to acknowledge the support of the previous directors of the programme Charles Desforges and John Kanefsky and the current director Andrew Pollard. I would also like to thank the School of Lifelong Learning in particular Trevor Corner and the School of Health & Social Science at Middlesex University for their support of the project.

I would also like to thank Professor Sally Glen (City University) and Professor Sharon Huttly (London School of Hygiene & Tropical Medicine) for their comments and suggestions on the report. I am however fully responsible for any errors or shortcomings .

Abstract

Introduction

Problem Based Learning is an approach to teaching and learning that has been designed using theory and research evidence about the nature of learning and of professional expertise and has been widely adopted. The advocates of Problem Based Learning claim many advantages for the approach. However it is often unclear what is meant by the term, and claims appear to be based on anecdotal evidence or small scale evaluative studies. There appear to be few reports of the use of Problem Based Learning in continuing professional education.

Methods

This report analyses a Problem Based Learning curriculum and evaluates its effectiveness in a continuing education programme for nurses in England. Using a randomised experimental design the learning outcomes of students who followed a 'traditional' curriculum were compared with the students who followed a Problem Based Learning curriculum in the same educational programme. The programme lasted one academic year and was undertaken on a part time basis. Five 'teachers' participated in the study all of whom were volunteers. The two teachers who facilitated the experimental Problem Based Learning groups undertook various staff development activities to prepare for their role as facilitators. The students were qualified nurses from five NHS hospitals who applied to take the programme during the study period. Thirty five students were allocated to the experimental (PBL) curriculum of whom 20 subsequently completed. Thirty four students were allocated to the control (SGL) curriculum of whom 31 subsequently completed. The students had no previous experience of Problem Based Learning.

Data Collection and analysis

A wide range of student outcomes were investigated using a variety of existing and new research instruments. Data on the process of curriculum development, programme delivery, students and teacher response were collected using non participant observation, teacher diaries, and researcher field notes. Qualitative data were analysed using the Framework method. Descriptive and Inferential Statistical analysis was carried out using SPSS. Analysis followed an estimation approach. Standardised effect sizes (*d*) with 95% confidence intervals were calculated to estimate the difference in outcomes between students in the experimental (PBL) and control (SGL) curricula.

Results

Students and teachers found it difficult to adapt and come to terms with the Problem Based Learning approach. It was apparent that Problem Based Learning did not meet the students normative expectations of 'teaching and learning'. Problem Based Learning appeared to cause the students great anxiety which lead to tensions between the teachers and students in the experimental (PBL) curriculum. This is reflected in the differences in student satisfaction indicators between the two curricula which all favoured the control (SGL) curriculum. Other

results showed that students in the control (SGL) curriculum were more likely to perceive that the educational programme had changed their practice. A greater proportion of students in the experimental (PBL) curriculum did not meet the threshold score indicating readiness for self-directed learning after completing the programme. The results exclude important statistically significant differences in impact between the two programmes on changes in approaches to learning, managers rating of performance, changes outside work, likelihood of taking on a teaching task in the workplace and on changing jobs.

Discussion and conclusion

The mixed results from this study appear to contradict those obtained in some other studies of Problem Based Learning. The results are consistent with other studies using experimental designs. Different forms of Problem Based Learning may produce different results, and quality iof implementation may also be an important factor. However, if such 'local' factors are important then it suggests that the context and culture in which Problem Based Learning is implemented is at least as important as the approach itself. The study found strong evidence of student dissatisfaction and of a disjunction between Professional, Student expectations and Problem Based Learning practice. It is argued that Problem Based Learning theory and practice lacks an adequate conceptualisation of the relationship between the different conceptions of teaching and learning held by different stakeholder groups in continuing professional education. Within the discourse of Problem Based Learning it appears to be taken for granted that everyone shares the principles, aims and values that underpin the approach i.e. there is a lack of recognition that pedagogy is a site for struggle between a number of competing discourses.

Part I

Introduction and methods

Part I Contents

	Page
Preface	**11**
Introduction	**12**
The context, continuing professional education in healthcare	12
Problem Based Learning an alternative model for continuing professional education?	12
What is not Problem Based Learning ?	14
An overview of key concepts in the theory and practice of Problem Based Learning	16
The aims of Problem Based Learning	17
Evidence about the effectiveness of Problem Based Learning	18
Study design and methods	**23**
Study research questions and methodological approach	23
Research design	26
Evaluating a complex intervention	30
Sample size	31
Recruitment and randomisation procedure	36
Selection of outcome measures and instrumentation	37
Reliability of assessment instruments	38
Validity of assessment instruments	38
Study evaluation framework	41
Framework Category Learner satisfaction	42
Learner satisfaction - Course Experience Questionnaire	42
Learner satisfaction – Nominal Group Techniques	43
Learner satisfaction – Telephone exit interviews	43
Framework category Skills, Personal and Propositional Knowledge - assignments	44
Minimising observer bias – External, independent blind marking	45
Improving the reliability and validity of the expert marking	46
Framework category Skills, Personal and Propositional Knowledge – Groupwork assessment	48
Framework category 'Approaches to Learning' - ASSIST	52
Longer term effects - Follow–up surveys	53
Pilot student questionnaire	55
Pilot line managers questionnaire	57
Pre-testing of pilot student follow-up questionnaires	58
Pre-testing of pilot line managers follow-up questionnaire	59
Development of 2nd version of questionnaires	60
Pre-testing of the 2nd version of the student follow-up questionnaire	62
Administration of the follow-up questionnaires	63
Timing of data collection	63
Economic evaluation	64
General Strategy for data analysis	65

Figures **Page**

Figure1.1 Summary of the key features and conceptual basis of Problem Based Learning 15
Figure 1.2: The SOLO taxonomy and hierarchy of verbs that indicate increasing cognitive complexity 47
Figure 1.3: Preliminary operationalisation of 'performance in practice' 55
Figure 1.4. Operationalisation of concepts in 2nd (final) student follow-up questionnaire 61

Boxes

Box 1.1 Areas of differentiation between Problem Based Learning and Traditional Curricula 17
Box 1.2 'Skills' that Problem Based Learning develops 18
Box 1.3: Sample size calculations for the study Randomisation procedure 35
Box 1.4 Microsoft Excel randomisation procedure 36
Box 1.5: Framework for the evaluation of continuing professional education 37
Box 1.6 Instructions for Nominal Group Technique 43
Box 1.7: Written assessments used on advanced diploma programme 44
Box 1.8: Marking process recommendations for external markers 48
Box1.9: Revisions to the follow-up questionnaires based on the systematic review of questionnaire design 60
Box 1.10: General Strategy for data analysis 65

Tables

Table1.1: Threats to statistical conclusion validity 28
Table 1.2: Threats to construct validity 29
Table 1.3: Threats to external validity 29
Table 1.4: Threats to Internal validity 30
Table 1.5: Previous module assessment results by assessment category 33
Table 1.6: Expected actual and cumulative proportions for effect size $d = 1.0$ 33
Table 1.7: Evaluation framework and summary of outcome measures and instrumentation 41
Table 1.8: Module assessment scores obtained on previous A24 & A25 programmes. 45
Table 1.9: Pilot student follow-up questionnaire: 'Self directed learning' performance items 56
Table 1.10: Pilot Student follow-up questionnaire: 'Clinical practice performance' items 56
Table 1.11: Pilot student follow-up questionnaire: 'Team member /leader performance' items 57
Table 1.12: Student follow-up questionnaire pilot - Cronbach's $\propto$ coefficients 59
Table 1.13: Student follow-up questionnaire pretesting of final version - Cronbach's $\propto$ coefficients 62
Table 1.14: data collection timetable module 1, 2001-2002 63

Preface: About this report

This is a report of a research study that aims to make a contribution to answering the question 'What forms of Problem Based Learning result in what outcomes for which students in which contexts?' It does so through evaluating the impact of a particular form of Problem Based Learning on a wide range of learning outcomes in a University run continuing education programme for nurses in England. What then can the reader hope to learn from this report? In the introduction to their report of a review on Teaching and Learning in the College Classroom, William Mckeachie and his colleagues addressed this question (Mckeachie et al, 1986). They argued that teachers should not expect to learn about entirely new ways of teaching or expect the broad generalisations that are the province of educational theorists. Similarly they should not expect to learn about *the* best way of teaching. Instead they point to a number of possible contributions that research may make. First is a more precise determination of the limits of generalisations. Second is disproof of faulty maxims and third is a better understanding of the way in which successful teaching strategies work. To which could be added a better understanding of why particular strategies do not appear to work in particular settings. Finally they argue that research affects the way we think about teaching and learning and in particular why and how we do what we do.

The report is divided into three parts. Part I sets the background for the project. A general overview of Problem Based Learning is given along with the methodological approach and methods used in the study. The design and implementation of the experimental (Problem Based Learning) and control (Small Group Learning) curricula are given in Part II. This part of the report seeks to ensure that there is no 'black box' of unknown practices at the heart of the study. A model of the 'inner teaching and learning environment' is used as a conceptual framework for this process. In an effort to integrate discussion of theory and practice, a detailed account of the theoretical basis of the constituent parts of Problem Based Learning is given alongside the description of their application in this study. Part II also includes analysis of the context of the study and of the responses of teachers and students. Part III contains the results, discussion and conclusions.

Problem Based Learning for continuing professional education: An exploration of the method and its effectiveness.

Introduction: The context, continuing professional education in healthcare

The improvement of health care quality requires continuing professional development amongst nurses and other health care professionals (Department of Health, 1999; Department of Health, 2001). Many references have been made to the importance and value of continuing professional education (CPE) for professional development and it is estimated that in the UK alone the National Health Service spends approximately £1billion per year on Continuing Professional Development (Brown et al, 2002). However, little evidence exists as to the actual impact that continuing education has on nurses ability to deliver higher standards of care (Barriball et al, 1992; Waddell, 1991; Wood, 1998). Research on the effectiveness of Continuing Professional Education across the health disciplines have raised questions about the effectiveness of these efforts concluding that the distribution of educational materials and formal Continuing Professional Education alone are ineffective strategies (Davis et al, 1999; NHS Centre for Reviews & Dissemination, 1999). This may in part be due to the pervasiveness of the 'update' model of Continuing Professional Education. In this model Professionals are removed from the workplace to a Higher Education institution for their knowledge to be 'updated' by an 'expert' in their respective field (Nowlen, 1988).

Problem Based Learning an alternative model for continuing professional education?

Problem Based Learning offers an alternative philosophy and method for Continuing Professional Education. It has been introduced into education in many professional fields including medicine, nursing, dentistry, social work, management, engineering and architecture. In its modern guise Problem Based Learning started to become a feature of educational programmes during the 1960's. Since then there has been a steady growth in the number of

programmes and institutions that have adopted Problem Based Learning around the world. The rise in the reported use of Problem Based Learning in medical education since its adoption by the new medical school at McMaster University in 1969, has been unsurpassed (Johnson & Finucane, 2000). By 1995 there were reports of the widespread use of Problem Based Learning in the USA, from Europe, The Middle East, The Far East and Australia (Bligh, 1995). It is not clear how many medical schools in the UK currently use a variant of Problem Based Learning but it is thought to be at least ten. It is not clear when or where Problem Based Learning first began to be used in Nursing Education programmes. Its adoption in Nursing Education appears to mirror the global reach of Problem Based Learning in Medical education, with reports of the use of Problem Based in nursing programmes emanating from Canada (Rideout et al, 2002), Australia (Alavi, 1995), the USA (White et al, 1999), the UK (Biley, 1999), China (Zhang & Zhang, 2000) and Egypt (Habib et al, 1999) to list but a few. However, it is not always clear what exactly is being done in the name of Problem Based Learning (Maudsley, 1999; Newman, 2003). There are also a growing number of references in the literature to 'adapted' or 'Hybrid' Problem Based Learning courses and courses called 'Enquiry' or 'Inquiry' Based learning which are apparently based on but not the same as Problem Based Learning (Cleverley, 2003; Margetson, 1998; Savin-Baden, 2000b).

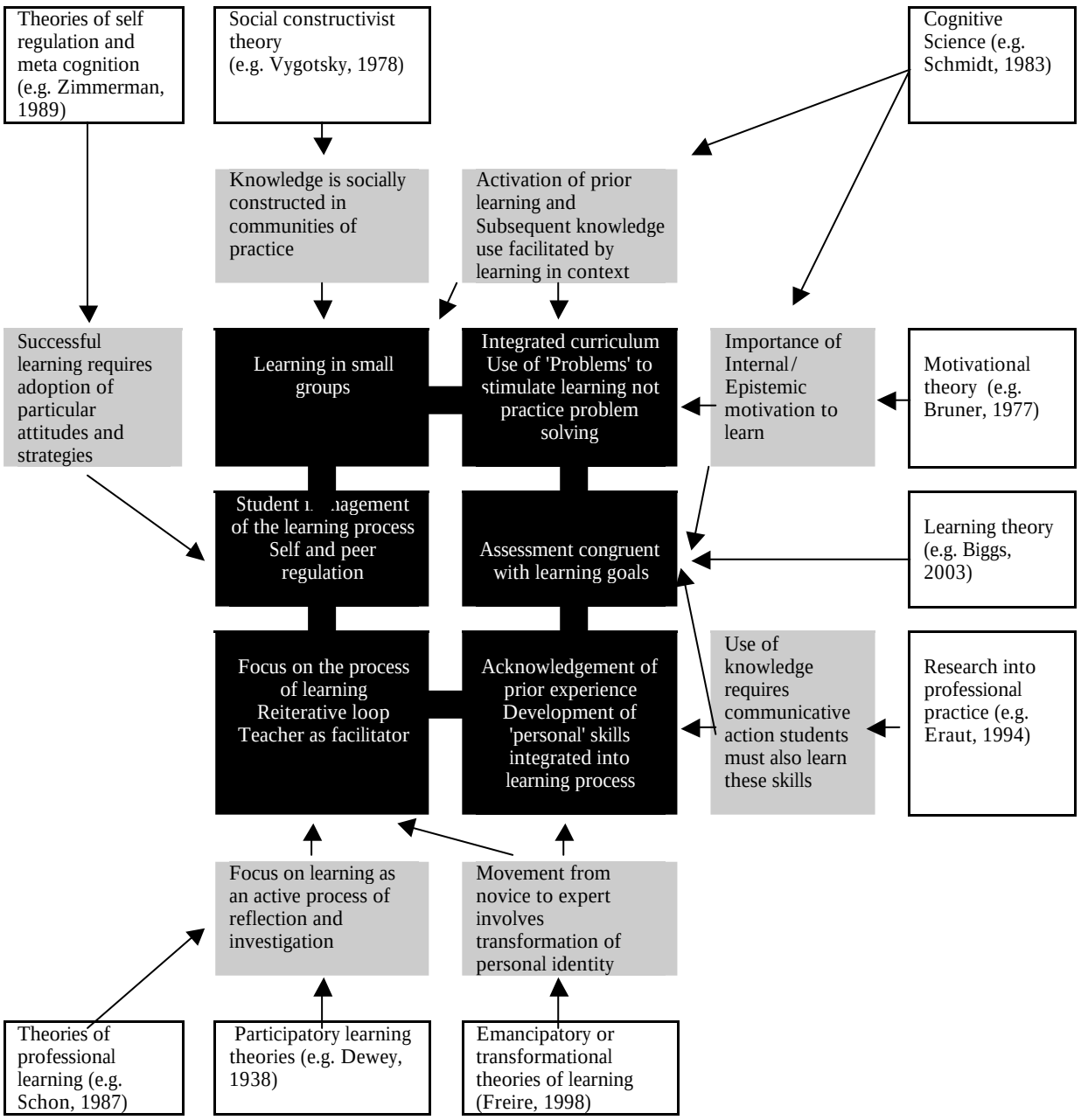
This transformation has been encouraged by an almost evangelical movement that has published of a wealth of anecdotal material extolling the virtues of Problem Based Learning (Wilkie, 2000). Problem Based Learning has been endorsed by a variety of national and international organizations. These include the Association American Medical Colleges (Muller, 1984), the World Federation of Medical Education (Walton & Matthews, 1989), The World Health Organization (World Health Organization, 1993), the World Bank (World Bank, 1993) and the English National Board for Nursing Midwifery and Health Visiting (English National Board, 1994). In recent years the advantages that are claimed for Problem Based Learning have become part of the generally articulated outcomes for education at all levels (Hmelo & Evenson, 2000).

What is not Problem Based Learning ?

There is no single unanimous position about the theoretical basis for, or practice of, Problem Based Learning. There is not even agreement about whether there is or should be one type of Problem Based Learning or many variants (Engel, 1991; Savin-Baden, 2000b). The wide dissemination of Problem Based Learning has 'de facto' spawned many variations (Barrows, 2000a). Maudsley (1999) argues that the label Problem Based Learning is often borrowed for prestige or subversion, adorning many narrowly focused single subject courses within traditional curricula that do not use Problem Based Learning at all. This claim would seem to be supported by the findings of a review of the curricula of American Medical Schools that claimed to use Problem Based Learning. This found that Problem Based Learning was being used as a generic category which included almost any teaching approach (Myers Kelson & Distlehorst, 2000).

An important distinction at the heart of Problem Based Learning is that with problem solving learning. Bereiter and Scardamalia (2000) distinguish between PBL (uppercase) and pbl (lowercase). Lowercase pbl refers to an indefinite range of educational approaches that give problems a central place in the learning activity. Whereas practitioners of 'PBL' uppercase tend to adhere to the structures and procedures first systematised by Howard Barrows (1986). Central to this system is a conception of learning as an integrated process of cognitive, metacognitive and personal development. The use of the term 'Problem' as in 'Problem Based Learning' does not therefore imply a foundationalist view of knowledge and learning (Margetson, 1993). The implication of foundationalism appears to be one of the main reasons why a number of nursing programmes have adopted the term 'Enquiry' or 'Inquiry' Based Learning (see for example Cleverley, 2003). Howard Barrows argues that a more accurate title for the model he and his collaborators developed might be "student-centered, problem based, inquiry-based, integrated, collaborative, reiterative, learning (Barrows, 2000b). However the label Problem Based Learning has stuck. This uppercase 'PBL' was the focus of this research.

Figure1.1 Summary of the key features and conceptual basis of Problem Based Learning



An overview of key concepts in the theory and practice of Problem Based Learning

It is important to distinguish the particular model of Problem Based Learning on offer in a programme. The model of Problem Based Learning used in this study is described in detail in part II of the report. Even within 'PBL' there is variation in the way that Problem Based Learning is theorised, described and practised. However, it is possible to identify what appear to be the key features of curricula that use Problem Based Learning and the concepts that underpin them. The philosophical and theoretical underpinnings of Problem Based Learning were not explicit in the early Problem Based Learning literature (Rideout & Carpio, 2001). Howard Barrows, a pioneer of Problem Based Learning, explains that the he and the other developers of the original the McMaster Problem Based Learning curriculum had no background in educational psychology or cognitive science. They just thought that learning in small groups through the use of clinical problems would make medical education more interesting and relevant for their students (Barrows, 2000a).

The key features and the concepts that underpin the approach are summarised in figure 1.1 above. The solid blocks at the centre summarise the key features, the grey shaded blocks the concepts underpinning each feature and the clear boxes the theoretical basis of the concepts. Each of the key features is discussed in more detail in Part II of the report. It is not the intention here to give a detailed account of the concepts and theories that underpin Problem Based Learning as the focus of the study is the approach itself rather than its underpinning theory. The development of Problem Based Learning in medical education appears to have been heavily influenced by Cognitive Science (Norman & Schmidt, 1992; Schmidt & Moust, 2000; Schmidt, 1983; Schmidt, 1993). But as Problem Based Learning has expanded into other disciplines interpretation of Problem Based Learning has expanded to include educational theories that emphasise other aspects teaching and learning such as participation (Dewey, 1938), Reflection (Schon, 1987) and the communal social construction of learning (Vygotsky, 1978).

All these 'theories' can be considered to be examples of constructivism whether social, cognitive or both (Tynjala, 1999).

Walton and Matthews (1989) argue that Problem Based Learning is to be understood as a general educational strategy rather than merely a teaching approach. They highlight three broad areas of differentiation between Problem Based Learning and the 'traditional' subject centered approaches (see box 1.1). Engel (1991) also focuses on curriculum design as a major area of difference. He describes the essential characteristics of problem-based curricula as cumulative (repeatedly reintroducing material at increasing depth), integrated (de-emphasising separate subjects), progressive (developing as students adapt) and consistent (supporting curricula aims through all its facets). Savin-Baden (2000b) identified five models of Problem Based Learning in operation in different curricula. She argues that the important differentiation is the way that knowledge, learning and the role of the student are conceptualised and manifest in the curriculum.

Box 1.1 Areas of differentiation between Problem Based Learning and Traditional Curricula (Walton & Matthews, 1989)

- Curricula Organisation: Around problems rather than disciplines, integrated, emphasis on cognitive skills as well as knowledge.
- Learning environment: use of small groups, tutorial instruction, active learning, student centered, independent study, use of relevant 'problems'.
- Outcomes: Focus on skills development and motivation, abilities for life long learning

The aims of Problem Based Learning

Dolmans and Schmidt (2000) give the aim of Problem Based Learning as helping students to develop rich cognitive models of the problems presented to them. Bailey and colleagues (2003) emphasise enhancing enculturation into clinical community of practice as an aim of Problem based Learning. Engel (1991) argues that where Problem Based Learning is adopted in professional education one of the aims is to assist students towards achieving a specific set of competencies, that will be important to them throughout their professional life, irrespective of

the profession in which they will come to practice. These are summarised in Box 1.2 along with other concepts/ skills developed by Problem Based Learning identified by Woods (1995).

Box 1.2 'Skills' that Problem Based Learning develops (Engel, 1991; Woods, 1995)

- Awareness (active listening)
- Personal learning preference
- Defining real problems (goals, mission, vision)
- Problem solving
- Learning skills (Laws, theories, concepts etc.)
- Look back and extending experience(recognising fundamentals in a given situation)
- Strategy (planning)
- Creativity
- Decision making
- Stress management
- Time management
- Group & chair person skills
- Managing change
- Interpersonal skills
- Coping creatively with conflict
- reasoning critically and creatively
- adopting a more universal or holistic approach
- practising empathy, appreciating the other person's point of view
- collaborating productively in groups or teams
- Identifying own strengths and weaknesses and undertaking appropriate remediation (self-directed learning)
- Self assessment
- Obtaining criteria
- Self - directed life time learning

Evidence about the effectiveness of Problem Based Learning

Problem Based Learning has arguably been one of the most scrutinised innovations in professional education (Maudsley, 1999). A simple illustration of this is that a search of the MEDLINE bibliographic database on-line via the PUBMED interface using the search terms 'Problem Based Learning' in Winter 2002/3 yielded a reference list of over 1000 citations. The literature on Problem Based Learning is spread over many different journals, books and databases and many subjects and disciplines. A brief search using the terms Problem Based Learning produced 804 'hits' on the Science Citation Index, and 384 in the Social Science Citation Index[1]. Woodward (1997) points out that empirical evidence that supports the theories that underpin Problem Based Learning is not the same as empirical evidence to support the claim that it produces practitioners with consistently high levels of performance that are maintained throughout their professional career. Block and Moore (1994) argue that despite the fact that many useful studies of problem Based Learning exist, selection bias and the absence of control groups limit the conclusions that can be drawn. Colliver (2000) re-ignited the debate

[1] February 2003 via WWW using Ovid interface

about the effectiveness of Problem Based Learning, and by implication the methods that should be used to investigate it, by arguing that many studies have erroneously claimed effects for Problem Based Learning when it was more likely that the effects were due to selection and philosophy of care differences.

In order to make the task of reviewing this literature more manageable this review of the evidence was initially confined to 'reviews' of the effectiveness of Problem Based Learning. There have been at least five 'reviews' of Problem Based Learning that have attempted to provide evidence about the conditions and contexts in which Problem Based Learning is more effective than other educational strategies. A major limitation of these reviews is that they include, with one or two exceptions, only studies of Problem Based Learning in the education of health professionals. Three of the reviews were published in the same journal in the same year (Albanese & Mitchell, 1993; Berkson, 1993; Vernon D.T & Blake, 1993). These three reviews, which are perhaps the most well known, are difficult to interpret due to the lack of clarity about the review methods used and apparent differences in approach between the reviews. The reviews include primary studies with different designs and of differing quality (Wolf, 1993). Of the citations identified by the review authors as providing 'evidence' about Problem Based Learning only eight appear in all three reviews, whereas 49 citations appear in only one out of the three.

The criteria for inclusion of studies in a 'Meta-analysis' of Problem Based Learning carried out by Van Den Bossche and colleagues (2000) are explicit. However the study design and quality criteria applied to the primary studies appear to be fairly minimal, raising the possibility that studies with significant weaknesses in terms of bias minimisation have been included in the review. The authors recognised the risk of bias in the location of studies and described, by the standards of most reviews, a reasonably comprehensive search strategy. However the search included only a limited number of Bibliographic Databases (not including MEDLINE) and the

search strategy only a limited number of terms and would therefore also appear to be inadequate in these respects (Egger & Smith, 1998).

Smits and colleagues (2002) carried out a review of the effectiveness of Problem Based Learning in continuing medical education. Their explicit search strategy included a wide range of bibliographic databases but it appears that limited attempts were made to locate the so-called 'grey' literature. This review adopted strict methodological inclusion criteria by including only randomised and controlled trials. Whilst this will have reduced the risk of bias in the individual studies (Cook & Campbell, 1979) it may also have meant that potentially useful studies of Problem Based Learning using other designs were excluded.

The reviews all provide only limited descriptive information about the educational interventions that are called Problem Based Learning or the interventions to which the Problem Based Learning is compared. Unsurprisingly the reviews referred to above came to differing conclusions. Vernon and Blake (1993) concluded "results generally support the superiority of the Problem Based Learning approach over more traditional academic methods". Albanese and Mitchell (1993) whilst acknowledging the weaknesses of the research literature concluded that Problem Based Learning was more nurturing and enjoyable and that Problem Based Learning graduates performed as well and sometimes better on clinical examinations and faculty evaluations. However, they also concluded that Problem Based Learning graduates showed potentially important gaps in their cognitive knowledge base, did not demonstrate expert reasoning patterns, and that Problem Based Learning was very costly. Berkson (1993) was unequivocal in her conclusion that "the graduate of Problem Based Learning is not distinguishable from his or her traditional counterpart". She further argued that the experience of Problem Based Learning can be stressful for the student and faculty and implementation may be unrealistically costly. The two more recent reviews also came to differing conclusions. Van Den Bossche and colleagues (2000) concluded that Problem Based Learning had a positive robust effect on the skills of students but a negative non-robust effect on knowledge. The

review by Smits and colleagues (2002) concluded that there was no consistent evidence that Problem Based Learning is superior to other educational strategies in improving doctors knowledge and performance. The reviews themselves therefore provide contradictory evidence about the effects of different kinds of Problem Based Learning in different learning contexts.

Educational theorists and researchers do not agree about the purpose and design of evaluative studies. This is mirrored within Problem Based Learning. Whilst experimental designs have been used there is disagreement about their value even within medical education (Johnson & Finucane, 2000; Torgerson, 2002). There are also disciplinary differences with apparently very few experimental studies of Problem Based Learning outside medical education. A more detailed consideration of this issue is given in the discussion of the rational for and design of this study in part II of the report. To obtain a clearer picture of what high quality research studies indicated about the effectiveness of Problem Based Learning, a research project involving secondary data analysis in the form of a 'review of reviews' was carried out as part of the Project on the Effectiveness of Problem Based Learning (Newman, 2003). Readers are advised to refer to the study report[2] for details of the methods and results of this study only a summary is given here.

Of the 90 studies identified as providing evidence of the effectiveness of Problem Based Learning in the five reviews referred to above, only 15 met the inclusion criteria for the 'review of reviews'. Three of these studies did not include any data in the reports seen. Not all of the outcomes reported in the 15 studies met the quality criteria. The details of the results of the included studies are discussed in a comparison with the results from this study in the discussion section of the report (part III) only a brief overview is given here. The results regarding cognitive development as measured by assessment of one kind or another varied. Of the 39 outcomes reported 16 favoured Problem Based Learning and 23 the control group. A pilot meta-analysis carried out as part of the review arrived at a mean effect size estimate of d = -0.3

[2] Report available from the project website http://www.hebes.mdx.ac.uk/teaching/Research/PEPBL/index.htm

i.e. in favour of the control group. However, this result should be treated with caution as the outcomes included are not independent (Hedges, 2003).

Only three of the included studies reported data that can be interpreted as measures of 'improvements in practice'. One study appeared to show that Problem Based Learning students held more desirable attitudes towards practice (Moore et al, 1994). Of the seven outcomes reported in a study of Problem Based Learning in a nursing programme two favoured the Problem Based Learning group (Lewis & Tamblyn, 1987). In a study of the use of Problem Based Learning to improve General Practice consultation skills only one outcome measure from nine favoured Problem Based Learning (Grol et al, 1989). Two of the included studies assessed changes in student learning styles and appeared to suggest that Problem Based Learning had a favourable impact on student learning styles (Coles, 1985; Moore et al, 1994). Whilst student satisfaction was reported in many of the studies considered in the review, in only one study did the measurement of this outcome meet the inclusion criteria. In this study, students in the Problem Based Learning curriculum appeared to rate their programme more highly (Moore et al, 1994).

In addition to the reviews highlighted above effort was also made to identify studies of the effectiveness of Problem Based Learning in nursing education and/or continuing education that were not included in the above reviews. The findings of the identified studies are discussed in detail in part III of the report in a comparison with the findings from this study.

Study design and methods

Study research question and methodological approach

The study can be located under the broad heading of evaluation research. The broad aim of evaluative studies of Problem Based Learning will be to find out what kinds of Problem Based Learning produce what learning outcomes for which students in which contexts and to ascertain the relative advantages offered by adopting the Problem Based Learning approach compared with any other. The research question in this study was 'Does the use of a Problem Based Learning curriculum in a continuing nursing education programme result in higher levels of student's attainment when compared to a 'traditional' curriculum?'. The question is thus one about a cause and effect relationship.

In her extensive study of the history and sociology of the pursuit of 'knowing' Oakley (2000) notes that experimental methods are largely rejected by social science today whereas they were once 'the' apparatus espoused by social scientists for investigating cause and effect relationships between social phenomena. It is often supposed that experimental methods are the preserve of the so called 'natural' sciences. However as Oakley (2000) points out prospective experimental studies have had a long, if chequered and often misunderstood history in social research. The SPECTR database (Social, Psychological and Educational Controlled Trials Register), an offshoot of the Cochrane initiative, contained over 10,500 randomised trials in 2000, identified with comparatively little resource and effort (Petrosino et al. 2000).

Oakley (2000) argues that criticism and rejection of the experimental approach in Social Science can be viewed as part of the ongoing 'paradigm' wars (i.e. between 'quantitative' and 'qualitative' approaches) that have been a feature of social science theorising since the advent of the discipline itself. The relative influence of any one set of arguments has varied over time, between disciplines and between countries but she suggests that three overlapping themes can

be discerned from the various critiques. One theme involves the positioning of different occupational groups around different perspectives on knowing and ways of knowing. A common component of such appeals is the rhetorical appeal to the 'special' 'person-centred' and/or 'complex' nature of the specific discipline. A second theme concentrates on the unscientific and unethical ways in which experimental studies have been carried out in the past, including the lack of consent and information given to participants. Thirdly cultural standards about masculinity and femininity have infused the framing of techniques of knowing. Science and its goal of objectivity through experiments are viewed as reciprocally linked to patriarchal structures that oppress women and other excluded groups.

These three themes can be seen in the list of specific criticisms of randomised experimental designs in social research summarised by Boruch (1997). That they are unfeasible in the real world; that they are expensive and time consuming; that other methods such as quasi-experimental designs and modelling will give the answer just as well; that randomisation is impossible/unethical; that such designs ignore other useful data; that results ignore differences of outcome within the experimental group; that they are one shot affairs that provide no useful information about how to improve the programmes being evaluated; that the results lack generalisability. Examples of all of the above critiques are common in Education (see for example Hammersley 2000) Nursing (see for example (Blomfield & Hardy 2000) and also the Problem Based Learning literature (Norman & Schmidt 2000, Dolmans 2003, Farrow & Norman 2003, Norman 2003).

The claim that education is more 'complex' than any other area of social science does not receive universal support (National Research Council 2002). Even if it were the case the education is more complex this would seem to suggest a greater need for randomised experiments when evaluating effectiveness rather than a wholesale rejection of the approach. Whilst uncontrolled, unethical experiments are clearly not justifiable these are not the same thing as well done, controlled experiments where, as in this case, the participants are required to give

their informed consent to participation in a study approved by a suitably constituted ethics committee. There does not appear to be good reason to view any particular research approach as more feminine, more ethical and thus more person centred for all research methods are intrusive to some degree and involve selection in one way or another (Oakley 2000). Research is difficult, expensive and time consuming, whichever methods are used. As the list of studies on the SPECTR database demonstrates randomised experiments in education are feasible.

Concerns about feasibility may disguise the collective misunderstanding of researchers, funders and participants about the operation of 'chance' in all kinds of research, and the ways in which various different methods minimise the effects of certain kinds of 'chance' and quantify others. There are certainly some situations in which randomised experimental designs are unethical and/or impossible (McKee et al. 2000; Thomson et al 2004). However, there does not seem to be a general rule that can be applied in all circumstances, rather the arguments need to be considered in relation to the specific research question and social setting of any investigation.

Arguments that experimental designs are not efficient i.e. do not make full or good use of the data appear to be more about the conduct of the research rather than specifically of the design itself. Again this criticism could apply to any research study that was not well conducted and/or was conducted with limited resources. The methods and conduct of this study are described in detail below and it is argued that the results demonstrate that full use of both quantitative and qualitative data was made to provide useful information about Problem Based Learning for researchers, policy makers and practitioners.

Claims that other research methods can provide the answer just as well can only be tested through empirical study. There is substantial empirical evidence that randomised experimental designs provide more conservative estimates of effect than observational or quasi experimental studies (Shultz et al 1995, Boruch 1997, Mckee et al 2000). Commentators such as Norman

(2003) and Dolmans (2003) argue that 'modelling' is a more efficient method of investigating the effectiveness of interventions such as Problem Based Learning because the variation between students within any particular programme is greater than the variance between students in different programmes. As Colliver (2003) has pointed out educationalists seem remarkably reluctant to draw the obvious conclusion from such results i.e. that the different programmes are not having differential effects. Quite aside from this point, it is difficult to respond to this claim as 'modelling' can and does mean quite different forms of analysis ranging from 'simple' linear regression to complex multi-level simultaneous equation modelling. One of the main limitations of this approach is whichever form of 'modelling' is used the analysis remains correlational and thus suggests only the possibility that that the observed relationship maybe causal (Wolfe 1985). The way in which the randomised experimental design used in this study meets the requirements for the attribution of 'causality' and attempts to control the various threats to validity is discussed in detail in the next section.

Research design

The most common form of causal explanation is based on four principles (Blaikie, 2000):

- There is a temporal order in which cause must precede effect
- There is association that requires that the two events occur together
- There is elimination of alternatives in order to be able to claim that the effect was due to the specified intervention and not something else.
- Causal relationships are made sense of in terms of broader theoretical ideas or assumptions.

In the context of this study the broader social scientific concept of causal mechanism as a set of conditions that when taken together produce an effect informs interpretation of the data (Selltiz et al, 1976). The section below that reports the design and methods used in the study demonstrates how the first three of these principles were met. The search for the broader meaning of these answers will include linking the data to that from other studies of Problem Based Learning.

The first three of these principles are primarily issues of internal validity and as such are 'managed' through the selection of the research design and the management of the research process. All possible threats to internal and external validity cannot be controlled in any one study, complex educational programs are implemented differently in various settings and are influenced by a host of political and social contexts. For these reasons smaller studies aimed at minimising bias (internal validity concerns) and random error (statistical validity concerns) are valuable in new or innovative educational programmes (such as Problem Based Learning) (Benson & Michael, 1982). The threats to the validity of studies identified by Cook and Campbell (1979) are given in columns one and two of tables 1.1 to 1.4. The third column in the tables summarises how the study design attempted to minimise each 'threat'. Not all threats to validity can be controlled in a single study. External validity in particular in particular is difficult to establish. This issue is discussed further in PART III of the report.

A randomised experimental research design was used. Evaluations of study designs have demonstrated that the well designed and executed randomised experiment is superior to any other design at minimising bias and random error and thus is considered most useful to demonstrate programme impact (Boruch & Wortman, 1979). The experiment is a particularly efficacious design for causal inference. Random assignment creates treatment groups that are initially comparable (in a probabilistic sense) on all subject attributes. It can then be reasonably be concluded that any final outcome differences are due to treatment effects alone, assuming that other possible threats to validity have been controlled (Tate, 1982). The pragmatic trial design used meant that the environment in which the experiment was conducted was kept as close as possible to normal educational practice. There is no placebo or sham intervention and all students who took the programme were included in the evaluation (Torgerson & Torgerson, 2001).

Table1.1: Threats to statistical conclusion validity (after Cook & Campbell, 1979)

Threats	Features	Study response
a) Low statistical power	Type II error increases when alpha is set low and sample is small	Sample size calculation done using α=0.05 β=80% for an Effect size d=1.
b) Violated statistical assumptions	All assumptions must be known and tested	Selection of appropriate statistical tests for analysis. Testing of data prior to analysis
c) Error rate	Increases unless adjustments are made with the number of mean differences possible to test on multiple dependent variables	Phase II exploratory trial
d) Reliability of measures	Low reliability indicates high standard errors	Use instruments with established reliability, measure reliability in new instruments, use of repeated measures, use of multiple measures for single outcome
e) Reliability of treatment	Treatment needs to be implemented in the same way person, site time	Small study, limited number of teachers, teachers involved in design of curriculum and study. Observation of teaching sessions.
f) Random irrelevancies in setting	Environmental effects which may cause or interact with treatment effects.	Students in different classrooms each session.
g) Random heterogeneity of respondents	Certain characteristics in subjects may be correlated with dependent variables	Randomisation to treatment groups post-hoc analysis of covariance

The disadvantage of the pragmatic trial approach is that there is greater variation making it harder to detect small effects. A number of modifications of the simple two group experimental design were considered to help offset this including 'matching subjects' (Robson, 1993), 'repeated measure' or 'cross over' designs (Louis et al, 1984), 'Single subject (A/B)' designs (Robson, 1993) and the 'two group pre and post – test' design (Robson, 1993).

However the way that recruitment to the programmes was organised meant that it was not possible to obtain any data about the participants prior to them starting the programmes. It was also felt unacceptable to ask students to complete any kind of assessment at the beginning of the programme. Given the part-time nature and short duration of the programme it was felt unlikely that the requirements for adequate duration of intervention and washout period required for crossover or single subject designs could be met (Senn, 1993).

Table 1.2: Threats to construct validity (after Cook & Campbell, 1979)

Threat	Features	Study response
a) Inadequate explication of the constructs	Poor definition of the constructs	Detailed description of PBL as planned and as delivered
b) Mono-operation bias	Measurement of single dependent variable	Multiple outcomes, multiple measures of each outcome
c) Mono-method bias	Measure dependent variable in one way	As above
d) Hypothesis guessing	Subjects try to guess researchers hypothesis and act in a way that they think the researcher wants them to act	Use of standardised 'objective' outcome measures'
e) Evaluation apprehension	Faking well to make results look good	Multiple outcome measures and assessments
f) Experimenter expectancies	Experimenter may bias study by their expectations when entering into and during study	Blind assessment of outcomes by external independent observer
g) Confounding levels of construct	All levels of construct are not fully implemented along a continuum .	Regular observation of teaching delivery throughout programme
h) Interaction of different treatments	Subjects are part of other treatments rather than of intended one.	Randomisation
i) Interaction of testing and treatment	Testing may facilitate or inhibit treatment influences	4 Group design not possible no pre-test
j) Restricted generalizability	The extent to which a construct can be generalized from one study to another	'Thick' description of sample, intervention and control

Table 1.3: Threats to external validity (after Cook & Campbell, 1979)

Threat	Features	Study response
a) Interaction of selection and treatment	Ability to generalise the treatment to persons beyond the group studied	Sample from 5 different NHS trusts, different grades etc. – describe baseline characteristics.
b) Interaction of setting and treatment	Ability to generalise to other settings beyond the one studied	Experimental design, nurses from more than one hospital,
c) Interaction of history and treatment	Ability to generalise the treatment to other times beyond the one studied	Sample at various points in post registration career. Unlikely that study will coincide with any other specific event that may affect outcomes

Table 1.4: Threats to Internal validity (after Cook & Campbell, 1979)

Threats	Features	Study response
a) History	Event external to treatment which may affect dependent variable	Experimental design with random allocation
b) Maturation	Biological and psychological changes in subjects which will affect their responses	Experimental design with random allocation
c) Testing	Effects of pre-test may alter responses on post-test regardless of treatment	2 group post test design
d) Instrumentation	Changes in instrumentation, raters or observers	Use of same instruments and observers throughout study
e) Statistical regression	Extreme scores tend to move to middle on post-testing regardless of treatment	Random allocation not selection based on pre-test
f) Selection	Differences in subjects prior to treatment	Random allocation – baseline comparison of groups
g) Mortality	Differential loss of subjects during study	Intention to treat analysis
h) Interaction of selection with maturation, history & Testing	Other characteristic of subjects mistaken for treatment effect on post-testing, differential effects in selection factors	Random allocation
i) Ambiguity about direction of causality	In studies conducted at one point in time, problem inferring direction of causality	Randomisation
j) diffusion /imitation of treatment	Treatment group share the conditions of their treatment with each other	Classes scheduled on different days - students from different workplaces
k) Compensatory equalisation of treatment	It is decided that everyone in experimental or comparison group receive the treatment that provides desirable goods and services	Different teachers in control and experimental groups
l) Demoralisation of respondents	Members of the group not receiving the treatment perceive they are inferior and give up	Motivations for both groups are to pass and gain certificate. Both groups treated identically in research terms. Emphasise that no approach is 'better'

Evaluating a complex intervention

As the design of the study progressed it became apparent that evidential claims about Problem Based Learning lacked both methodological and conceptual clarity (Colliver, 2000; Maudsley, 1999). Furthermore, Problem Based Learning can be considered to be a complex intervention and thus subject to the specific difficulties in defining, developing, documenting and reproducing all such interventions. The Medical Research Council (MRC) framework for the design and evaluation of complex interventions to improve health (Campbell et al, 2000), is equally applicable to complex interventions in other fields such as education. The framework

utilises a sequential phased approach to the development of randomised trials of complex interventions. Using this framework this study can be considered a phase II exploratory trial. A phase II exploratory trial is concerned with defining the control intervention, estimating the size of the effect, identifying and piloting various outcomes and outcome measures.

Whilst the distinction between exploratory and definitive trials provides a useful framework for study design, in practice the boundaries between an exploratory (phase II) trial and a definitive (phase III) trial are blurred. In this study effect sizes and outcomes were identified prior to the study and thus are amenable to hypothesis testing. However given the notable difficulties in measuring the impact of education (Van Der Vleuten, 1996) and the lack of valid reliable instruments in Problem Based Learning, few of the instruments used in the study have been used in studies of the effectiveness of Problem Based Learning before. This practical blurring of the boundaries also highlights the conceptual blur between the two phases. Given the variety of educational contexts it is questionable whether there could be *a*' definitive trial of Problem Based Learning. It maybe that there will need to be definitive trials of Problem Based Learning in different education contexts of which Continuing Professional Education (CPE) is one.

Sample size

A detailed analysis of the beliefs, values and other relevant characteristics of the teachers and students who participated in the study are given in part II of the report. Sample size calculations were undertaken to estimate the sample size required to achieve a particular level statistical power and precision (Du Florey, 1993). The sample size required depends on four factors: Variance of the variable being studied; size of the effect of interest; Level of Significance; Power of the test. The choice of the level of significance and acceptable power are matters of convention (Altman, 1991). Generally speaking the larger the sample the smaller the effect size that can be detected. However, the question of what effect size is 'important' i.e. is

likely to be indicative of a 'real' change in outcome for a reasonable number of students is controversial.

The traditional approach to calculating effect size uses a measure of the mean divided by the standard deviation which gives a measure of effect size denoted by the use of the character '*d*' (Coe, 2002). Researchers in education and other fields continue to debate the practical significance of an effect size. A general recommended guideline across disciplines is that $d=0.2$ (small effect), $d=0.50$ (moderate effect), and $d=0.80$ (large effect) (Cohen, 1988). Given the problems of measurement error, non-response and the fact that the instruments are measuring latent variables it could reasonably be argued small or even moderate effects could be accounted for by such study artefacts (Gorard et al, 2002). At the other extreme it has been argued that an effect size of $d=2.0$ should be required where wholesale curriculum and organisational change is implied (Bloom, 1984). Colliver (2000) argued that $d=1.0$ should be the minimum size of effect required to justify the considerable change that Problem Based Learning requires.

A variety of sources could be used to supply the data required for sample size calculations. In the absence of any robust data from studies of Problem Based Learning with this particular professional group, marks obtained by students for assignments carried out for the same programme in previous years was used. These course assignments were marked using the standard University wide 20-point scale. Using records from previous courses details of assessment scores for the three course modules for 30 students were obtained. Students for whom all three-module scores could not be found were excluded. The mean score for the 90 modules was 8 with a standard deviation of 4.1

Table 1.5: Previous module assessment results by assessment category

University score Category	N	Score Category	%	Cumulative category	Cum %
1-3	14	CA3	15.6	CAC4	100
4-7	30	CA3	33.3	CAC3	84.5
8-12	35	CA2	38.9	CAC2	51.2
13-20	11	CA1	12.3	CAC1	12.3

There are numerous formulas for the calculation of sample size. The method selected depends on the desired level of precision required in the estimate, the size of the groups being compared, and the type of outcome measures being used. A specific formula for calculating sample sizes for ordered categorical data was used. The formula and calculations used in this study are given in box 1.3. The method uses the odds ratio to specify the effect size or proportion expected in each category. The odds ratio is the chance of a subject being in a given category or lower in one group compared with the other. The odds ratio was calculated from the previous course results.

If the number of categories is large it is difficult to postulate the number of people who would fall in a given category. For the purpose of sample size calculation there is little increase in power (and thus reduction in the number of subjects required) to be gained by increasing the number of categories beyond five (Campbell et al, 1995). The university 20-point scale was aggregated into five sub categories based on the assessment criteria. Given the very small proportion of previous assessment results that fell into the clear failure category (17-20) the lower two categories of results were combined.

Table 1.6: Expected actual and cumulative proportions for effect size d = 1.0 (based on previous assessment scores

University score Category	N	Actual category	Actual %	Cumulative Category	Cumulative %
1-3	14	CB4	32	CBC4	100
4-7	30	CB3	38	CBC3	68
8-12	35	CB2	23.4	CBC2	30
13-20	11	CB1	6.6	CBC1	6.6

The figures used the calculations in box 1.3 are based on identifying an effect size of d=1.0 (51%) reduction in categories one two and three (Mean of previous results = 8 standard deviation = 4.1 i.e. approx. 51%). CA = observed category CB = expected category. Table 1.6 shows the expected proportions in each category for an effect size of d=1.0 Using this approach the estimated sample size required to detect an effect size of d=1.0 with a power of 80% and 95% significance was 23 per group. It should be noted that that such a figure provides an 'estimate'. Such estimates are heavily dependent on the assumptions made to select the data that is used in their calculation. Post – Hoc power analysis i.e. based on the final sample size for each outcome was also carried out.

Box 1.3: Sample size calculations for the study (Campbell et al, 1995)

If the effect size is specified as an odds ratio:
$N = 6(z1-\alpha/2) + z1-\beta^2 / (\log OR)^2 / [1-(\Sigma)i=1^k Pi^{-3}]$

For a 2 sided $\alpha = 0.05$ then $z1-(\alpha/2) = 1.96$
For $\beta = 0.2$ then $z1-\beta = 0.84$
K = Number of categories
Pi = mean proportion in each category

Step 1- Odds ratio for each category

CA 1
OR ={CA1/(1-CA1)}/{CB1/(1-CB1)}
OR ={0.123/(1-0.123)}/{0.0615/(1-0.0615)}
OR ={0.123/0.877}/{0.0615/0.9385}
OR =0.14/0.07
OR =2

CA 2
OR ={CA2/(1-CA2)}/{CB2/(1-CB2)}
OR ={0.512 /0.488}/{0.261/0.739}
OR= 1.049/0.353
OR=2.97

CA 3
OR ={CA3/(1-CA3)}/{CB3/(1-CB3)}
OR ={0.845/0.155}/{0.431/0.569}
OR =5.45/0.76
OR =7.17

Step 2 Calculate the expected cumulative %

CBC1 = CAC1/{CAC1+OR (1-CAC1)} where CBC = expected cumulative category.

CBC1

CBC1 = 0.123/{0.123+2 (1-0.123)}
CBC1 = 0.123/{2.123 * 0.877}
CBC1 = 0.123/1.86
CBC1 = 0.066

CBC2
CBC2 = 0.512/{0.512+2.97 (1-0.512)}
CBC2 = 0.512/ 3.482* 0.488
CBC2 = 0.512/1.7
CBC2 = 0.30

CBC3
CBC3 = 0.845/{0.845+7.17 (1-0.845)}
CBC3 = 0.845/ 8.015*0.155
CBC3 = 0.845/1.24
CBC3 = 0.68

Step 3: Calculate the value of $(1-\{\Sigma\}p^3)$

Where Pi = mean actual % in each category = (CA+CB)/2

Pi1- 0.09 = (0.123 + 0.066)/2
Pi2- 0.31 = (0.389 + 0.234)/2
Pi3- 0.36 = (0.333 + 0.38)/2
Pi4- 0.24 = (0.156 +0.32)/2

$1-\{\Sigma\}p^3 = 1- (0.09^3+0.31^3+0.36^3+0.24^3)$
$1-\{\Sigma\}p^3 = 1- (0.0007+0.0298+0.0466+0.0138)$
$1-\{\Sigma\}p^3 = 1- 0.09$
$1-\{\Sigma\}p^3 = 0.91$

Step 4. Apply sample size numerator and correction factor
The numerator used for calculating the sample size is taken from table III in the paper by Campbell et al. Using the average odds ratio (4.04), for an 80% power and two sided 5% significance the numerator = 24.5.
24.5 *0.91 = 22 per group.

Applying the correction factor used when there are less than 5 categories (1.067) gives a total of 23 per group.

Recruitment and randomisation procedure

The ethics committee of the department concerned approved the study design and protocol. Students applied for the programme in the usual manner over a period of several months. After students had applied for the programme they were contacted by the Principal Investigator who met with them to explain the research. Written information was also provided. Participants were allocated to either the experimental (PBL) or control (SGL) curriculum group after they had signed a consent form to participate in the study. It was also necessary to inform participants which group they had been randomised to as early as possible in order to facilitate their release from the workplace on the appropriate days. To ensure that the size of the intervention and control groups were reasonably matched a block randomisation procedure (using blocks of 10) stratified by programme (i.e. medical /surgical) was used.

A unique study number was generated for each participant in the study. The Principal Investigator placed these numbers into sealed opaque envelope. The numbers were allocated to either the experimental (PBL) curriculum or the control (SGL) curriculum using the 'Random' function in Microsoft Excel as described in Box 1.4 below. The Principal Investigator offered each participant a choice of sealed envelopes. The number inside the envelope was unknown to either the Principal Investigator or the participant. The candidate opened the sealed envelope to reveal their identity number that was then checked against the randomisation list and the student informed which group they had been allocated to.

Box 1.4 Microsoft Excel randomisation procedure

1. In the first column of a worksheet sheet 20 cases were listed in the following sequence, PBL, SGL, PBL, SGL……
2. In the second column a number corresponding to the block was placed in each cell. The first 10 cases were labelled block 1 and the second 10 cases block 2
3. In the third column a random number was generated in each cell using the RAND() function. This inserts a random number between 0and 1 in each cell.
4. The fourth column contains the identity number that is unique to each case.
5. The first column is then sorted in ascending order by block and random number. The numerical order of the fourth column remained unchanged. Thus each identity number is randomly allocated to either the experimental or control groups.

Selection of outcome measures and instrumentation

Cervero's (1988) framework for the evaluation of continuing education for professionals was used as an initial guide the selection of appropriate outcome measures and instrumentation for the study (see box 1.5). The category 'Impact of the application of learning' refers to so-called 'second order' effects. In the context of this study this refers to whether there are measurable improvements in patient outcomes as a result of nurses undertaking a continuing nursing education programme. Measurement of such effects was beyond the scope of this study.

Box 1.5: Framework for the evaluation of continuing professional education (Cervero, 1988)

- Programme design and implementation
- Learner participation
- Learner satisfaction
- Learner knowledge, skills and attitudes
- Application of learning after the programme
- Impact of application of learning (second order effects – e.g. improvements in the health of patients)

Despite the extensive literature on assessment of professional competence there is little consensus about what exactly should be measured let alone how it should be measured (Van Der Vleuten, 1996). An important aspect of Problem Based Learning philosophy is the recognition of the fact that assessment has a major impact on learning. However, there is not a consensus on either the outcomes or methods of measurement that should to be used to evaluate the effects of Problem Based Learning on student knowledge, skills and attitudes. A range of student capabilities under this heading can be identified in the Problem Based Learning literature.

Given the methodological approach of the study and the limited time and resources available effort was made to identify existing sensitive, valid and reliable outcomes and instruments for which high levels of response could be obtained. The setting of the experiment i.e. as a pragmatic trial in a 'real world' education setting provided an additional set of constraints. Any research measurement needed to place as little burden on the students and teachers as possible and not to divert students from learning. It was therefore agreed that it would be unreasonable and impractical to require students to undertake any additional form of summative testing or

assessment. The selection and use of measurement tools for the study involved a trade off between reliability, validity, educational impact, acceptability and cost which are discussed in detail below.

Reliability of assessment instruments

The key problem identified in research on outcome assessment is the variability of candidate performance on even very similar cognitive tasks. This occurs whatever the competence being measured and whatever response format is used (with the possible exception of Multiple Choice Questions containing a large sample of items), suggesting that assessments containing a small sample of items e.g. essays, produce unstable or unreliable scores (Van Der Vleuten, 1996). Van Der Vleuten (1996) argues that the practical consequences of this are that the sample size of test items should be sufficiently large and the test designed such that the affect of variability on the precision of the instrument is minimised. Where possible each group of outcomes investigated in the study were measured using a variety of indicators and instruments.

Validity of assessment instruments

The assessment of validity i.e. that tests measure what they are required to measure, requires the identification of good criteria or standards. In most areas of professional competence good criteria and perfect standards do not exist (Van Der Vleuten, 1996). Problem Based Learning is no exception. A recent evaluation of the Problem Based BSc Nursing programme at McMaster University in Canada included use of 'The California Critical Thinking Skills Test' (CCTST) (Facione, 1990), (*Personal communication Professor Liz Rideout- McMaster University*). The CCTST is based on the consensus view of the critical thinker produced by the American Philosophical Association and has undergone extensive testing by the authors (Howell Adams et al, 1996). The California Critical Thinking Disposition Inventory (CCTDI) is a separate but related test of the

disposition to think critically. The CCTDI is based on the argument that the critical thinker also has a particular characterological profile, which disposes them to think critically (Facione et al, 1995), and has also undergone rigorous testing (Facione et al, 1994). Numerous criticisms have been made of both the CCTST and the CCTDI but they are probably as useful as any other standardised critical thinking test (Howell Adams et al, 1996).

However, the main problem of all such tests lies in the way that critical thinking is conceptualised independently of context. Fisher and Scriven (1997) argue that critical thinking is underpinned by informal logic, and is thus context dependent. Problem Based Learning is based on principles derived from cognitive psychology i.e. that knowledge is structured in semantic networks. Problem Based Learning scenarios create a semantic structure for the learning of knowledge which is similar to the semantic structure in which the knowledge will be applied thus enabling the recall of required knowledge (Gijselaers, 1996). It would therefore seem inappropriate to use context free critical thinking tests to measure outcomes achieved by Problem Based Learning.

Another 'validity' issue in relation to Problem Based Learning is the shared view amongst Problem Based Learning advocates that assessment drives learning. However, the consequences of this view are interpreted differently. Some writers suggest that both the response format and the content of the test must be appropriate to Problem Based Learning (Marks-Maran & Gail Thomas, 2000). Others argue that response format is of less consequence than content and test-design (Norman, 1991). The Multiple-Choice Question format was introduced to cope with the increased logistical demand for educational testing and to provide reliable assessment of student performance. Multiple Choice Questions have often been rejected for use in Problem Based Learning programmes for various reasons including the belief that they are only suitable to measure lower levels of taxonomic cognitive functioning (Van Der Vleuten, 1996). However others argue that there is no reason why Multiple Choice Questions cannot be used in Problem Based Learning assessment as the key issue is the quality of the design and administration of the test rather than the method itself (Swanson et al, 1991). The 'Progress test' (Arnold & Willoughby, 1990), used with slight variation in the Problem Based Learning programmes in a number of

Medical Schools including McMaster (Canada)(Blake et al, 1996), uses the Multiple Choice Question Format.

A number of assessment formats are claimed to provide a more valid measure of the learning developed by Problem Based Learning programmes. Modified Essay Questions (MEQ) have been used to assess Problem Based Learning in both clinical and pre-clinical courses. It is argued that the properly designed evolving Modified Essay Question opens up possibilities for exercising 'intelligent guessing' that mirrors the realities of clinical work and can thus measure abilities and attitudes that other assessment methods cannot (Knox, 1980). Although the reliability of the Modified Essay Question method has been established (Feletti, 1980), caution has been expressed about its misuse and over use in Problem Based Learning programmes (Feletti & Smith, 1986). Studies have also suggested that the Modified Essay Question measures nothing different from the Multiple Choice Question (Norman, 1989). Modified Essay Questions are used as part of the assessment programme on the BSc Nursing Programmes at Thames Valley University and the University of Dundee which both use Problem Based Learning (Marks-Maran & Gail Thomas, 2000; Wilkie, 2000). However, the reliability of these Modified Essay Questions has not been established. This and practical constraints prevented their use in this study.

The Triple Jump Exercise is a learning process measure widely used as an assessment tool in Problem Based Learning programmes (Painvin et al, 1979). The Triple Jump Exercise consists of three steps (jumps.) A structured oral examination based on one or more patient problems, a time limited study assignment in relation to the patient problems in the first oral and a repeat oral examination in which the quality of self – learning around the assigned topic is assessed. The Triple Jump Exercise is currently used in a number of Problem Based Learning programmes around the world, including the Problem-based BSc Nursing programme at McMaster University in Canada. The Triple Jump Exercise is a very time consuming, costly method of assessment with poor measurement characteristics (Blake et al, 1995). These factors combined with practical constraints prevented the use of the Triple Jump Exercise in this study.

Study evaluation framework

In order to further improve the validity of the outcomes and instrumentation used a summary of the claims made for Problem Based Learning produced by Engel (1991) was also used to guide the selection of appropriate outcome measures and instruments. However the group of measures and instruments eventually used did not fall neatly into either Engel's or Cerveros' categories. This is a reflection of both the conceptual and practical difficulties of assessment in this area. The student assignments for example could be argued to assess, cognitive, metacognitive, personal and propositional skills and knowledge. The final list of outcomes and instruments and the revised evaluation framework shown are shown in table 1.7 below. 'Study contexts, participants, curriculum theory and practice' are described in Part II of the report.

Table 1.7: Evaluation framework and summary of outcome measures and instrumentation

Cervero Category	PEPBL Evaluation framework	Measure
Programme design and implementation	Study contexts, participants, curriculum theory and practice (Described in Part II of report)	Tutor record of session content and activity Interaction analysis Non participant observation Tutor records of student attendance activity Interaction analysis Student study workload (self reported)
Learner participation Learner/teacher satisfaction	Learner/teacher satisfaction	Course Evaluation Questionnaire Observations Teachers Diaries Nominal Group Technique Drop-out rates Exit Interviews Students Follow-up questionnaire
Learner Outcomes	Skills, personal and propositional knowledge	Follow-up questionnaire of students Follow-up questionnaire of students' managers Assignments x 3 Approaches to Study Inventory (ASSIST) Self-Directed Learning Readiness Scale Group work video assessment
Application of learning after the programme	Approaches to learning	

Framework Category Learner satisfaction

It is often claimed that Problem Based Learning leads to increased levels of learner satisfaction or that students like Problem Based Learning (Wilkie, 2000). This would seem to be an important outcome both for its own sake and because of an imputed link between enjoyment, motivation and performance (Mckeachie et al, 1986). There are a number of ways of conceptualising enjoyment and satisfaction in an educational context and therefore a 'basket of indicators' approach was adopted. In this approach the same outcome is 'measured' using a variety of approaches/ instruments. The satisfaction scales included in the follow-up survey are discussed in the section on the follow-up survey.

Learner satisfaction - Course Experience Questionnaire

The Course Experience Questionnaire (CEQ) (Ramsden, 1992) is based on empirical and theoretical work on the quality of teaching in higher education. Students are asked to rate the quality of their programme using questions with a five point likert scale. The assessment contains five scales; teaching, goals, workload, assessment and student independence. The Course Experience Questionnaire was tested in 50 Australian education institutions on 4500 students cross a range of disciplines and was found to discriminate between teaching styles and quality within and between different subject areas (Ramsden, 1992). The use of the Course Experience Questionnaire is now compulsory in Australian Higher Education Institutions (Long & Johnson, 1997). The Course Experience Questionnaire was also used to evaluate student satisfaction on the Problem Based Learning Programmes in the Health Science Faculty at Griffith University in Brisbane (Margetson, 1995). The Course Experience Questionnaire has been updated several times. One reason for using the original version of the Course Experience Questionnaire is that the scale 'Emphasis on independence' (which appears very relevant to evaluations of Problem Based Learning) has been dropped from more recent versions of the scale now in widespread use.

Learner satisfaction – Nominal Group Techniques

There are a variety of Nominal Group Techniques (NGT) that can be used for problem identification (Delbecq & Van den Ven, 1971). The approach used in the study was a variation of the RAND form of Nominal Group technique (Black et al, 1998). The instructions given to students are shown in box 1.6. The Nominal Group technique was undertaken on the final day of each group's programme. The instructions were given to the students by the Principal Investigator. The Principal Investigator and the teacher left the classroom until the students had completed the exercise. After the students had completed the exercise the lists generated by the students were discussed with them to gain greater clarification.

Box1.6: Instructions for Nominal Group Technique

1. List five things that you have enjoyed about the programme
2. List five things that you found difficult on or about the programme
3. After all the group has completed parts one & two compile a group list using the items highlighted by each individual eliminating any duplications
4. Each member of the group has five points to award to the things that they enjoyed most from the group list. You can allocate the points in any way that you choose. For example you could allocate all points to one item or three points to one and two to another or one point to each of five different items. You do not have to give the points to the items that you chose originally, if you feel that there are other items on the group list that are more important.
5. Each member of the group has five points to award to the things that they enjoyed least from the group list (five = least enjoyable). You can allocate the points in any way that you choose. For example you could allocate all points to one item or three points to one and two to another or one point to each of five different items. You do not have to give the points to the items that you chose originally, if you feel that there are other items on the group list that are more important.
6. Add up the points on the list to arrive at five best and five worst things on the course. From the perspective of the group

Learner satisfaction – Telephone exit interviews

Telephone exit interviews were carried out with all students who discontinued the programme for whatever reason. The interview schedule was designed specifically for this study. Students were contacted as soon as the Principal Investigator became aware that they had left the

programme. The period of time between the students last teaching session and when they were contacted varied as it was often not confirmed for some weeks that a student had actually quit the programme as opposed to just being absent. The Principal Investigator contacted the student to arrange a convenient time for the telephone interview. During the interview the Principal Investigator made note of the student's responses and wrote up the interview immediately after the interview was complete. Analysis of the exit interviews was carried out by the Principal Investigator and comprised of reviewing the completed exit interview schedules to identify areas of commonality and difference in the students' accounts.

Framework category Skills, Personal and Propositional Knowledge - assignments

The written assessment methods currently used in both programmes use the free response format (see box 1.7). With their emphasis on self selection of topic, self-directed information searching and presentation of data in a clear focussed manner, written assignments are viewed as a relevant evaluation method within the Problem Based Learning approach (Rideout, 2001). They are widely used in assessment programmes on Problem Based Learning courses (Marks-Maran & Gail Thomas, 2000). The assignments are discussed in more detail in part II of the report. The pre-existing course assignments were congruent with the aims of Problem Based Learning and had the advantage that the students would be motivated to complete the assignments well given that they are a programme requirement. It was therefore decided that students' assignment scores should be used as one of the outcome measures for the research study.

Box 1.7: Written assessments used on advanced diploma programme

- Literature review and seminar presentation;
- Care study and supporting essay;
- Learning contract and reflective account

However, the poor intra and inter observer reliability of marker evaluations of free response assessments are well documented (Biggs, 1999; Brown et al, 1997; Swanson et al, 1991; Van Der

Vleuten, 1996). Analysis of available data on assignment scores from previous years of the participating programme reveal a skewed distribution towards the higher end of the marking scale which did not match the teachers verbal accounts of the performance of previous students (see table 1.8). It can be argued that the cause of these validity and reliability problems is the tutors marking rather than anything inherent in the method itself (Swanson et al, 1991). The provision of simple protocols to structure and score examinations has been demonstrated to significantly improve reliability as compared to free judgement (Verma et al, 1997).

Table 1.8: Module assessment scores obtained on previous A24 & A25 programmes.

Grades	% (N=90 modules*)
1-3	15.6
4-7	33.3
8-12	38.9
13-16	6.7
17-20	5.6

*30 participants x 3 modules analysed. Grade 1 highest.

Minimising observer bias – External, independent blind marking

There is evidence that unblinded outcome assessment, particularly for subjective outcomes (such as used here), is demonstrably associated with bias (Prescott et al, 1998). The assignment scores used for the research were therefore generated independently from the marks given by the programme teachers to meet the programme assessment requirements. Three nurse teachers from other UK universities were recruited to mark the assignments for the purpose of the research grade. Each marker was a nurse educator and had experience of teaching and marking in pre and post registration programmes. The markers had no previous connection with either Middlesex University or any member of the teaching or research team in the study. The markers were paid the standard University external examiner fee. The scripts were anonymized by removal of all identification except a student number, and sent to the external examiner by post for marking. The marking for research purposes was therefore carried out by independent experts, 'blind' to the allocation status of the students.

Improving the reliability and validity of the expert marking

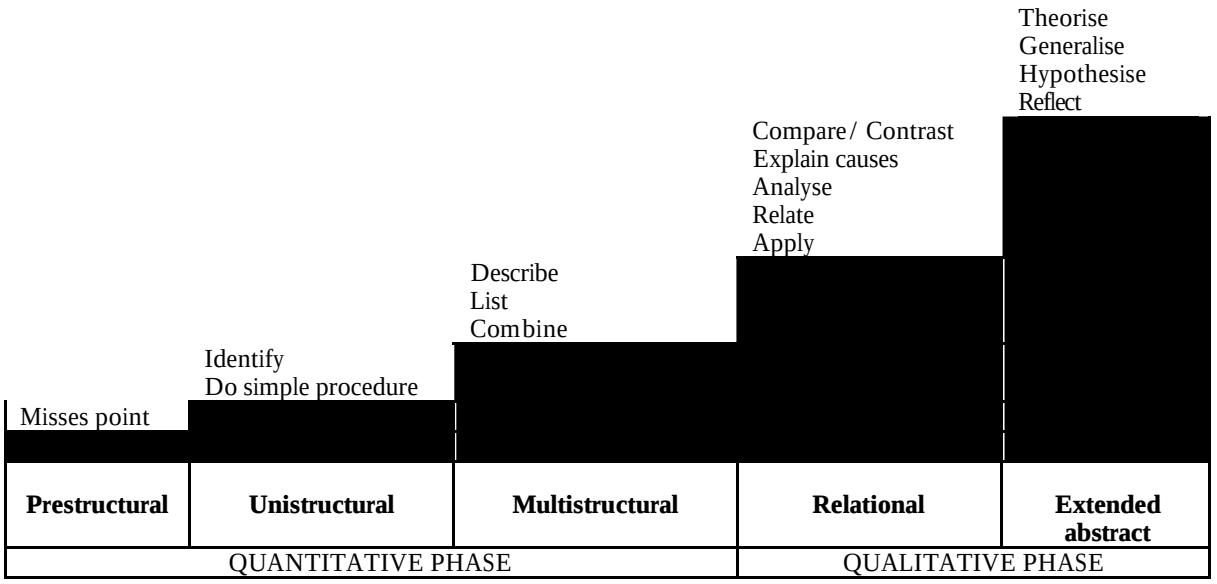
Despite agreement that marking protocols are useful there are huge variations in the types of protocol used and disagreement about the nature of the criteria that should be included. According to Biggs (1999) this is partly due to different views about 'learning' and assessment and also because of the dominance in Higher Education of the norm referenced approach to assessment. He argues that this often results in marking protocols that do not reflect what it is the 'teaching' is trying to achieve, either through omission or through the use of an analytic approach in which the big picture of performance is somehow lost in the detailed criteria. Detailed criteria have been shown to yield more to low level learning i.e. students can obtain high marks even though only lower level learning has been demonstrated. They also fail to improve reliability due to their difficulty in use (Brown et al, 1997). However, more detailed criteria can be useful for research purposes, but only to the extent that markers will actually use them.

The purpose of the programme assignments is to measure the extent to which a student has achieved the objectives or learning outcomes of the programme. The aim for the new protocols was firstly to ensure that what is marked reflects the programme objectives i.e. is valid. With respect to this point it should be noted that it was not the intention to develop new or different criteria that did not reflect the course objectives or the information that students were given. This would be of questionable validity. Secondly, to improve reliability i.e. the likelihood that the same person would make the same judgement about the same performance on two different occasions (intra-observer reliability) and different judges would make the same judgement about the same performance on the same occasion (inter-observer reliability).

In relation to validity the issue is to ensure that understanding is defined in ways that do justice to the topic/ content taught and level of study as exemplified in the in programme objectives (Biggs, 1999). The SOLO taxonomy provides a general framework for structuring levels of understanding. It is based on the study of student outcomes in a variety of academic content areas which demonstrated that as students grow the outcomes of their learning display similar

stages of increased structural complexity. Levels of understanding can be described as verbs in ascending order of cognitive complexity that parallel the SOLO taxonomy (see fig1.2) (Biggs, 1999).

Figure 1.2: The SOLO taxonomy and hierarchy of verbs that indicate increasing cognitive complexity



An analysis of the programme objectives and assignment information given to students (see part II of the project report) identified that most of the verbs used are firmly in the relational stage of the taxonomy extending in some parts to the extended abstract level. The purpose of the assessments as stated in the student handbook is given as "to reveal the student's ability to synthesise and evaluate the theoretical issues of each of the modules and to facilitate student's exploration of their value system which underpins their professional practice". The requirement for this level of understanding is congruent with the final year undergraduate, status of the programme. The marking protocol improves reliability by identifying clearly and unambiguously what the marker should be looking for in terms of level of understanding displayed in the students writing and how these components should be weighted when considering the overall mark allocated.

The new marking protocols were based on existing standard models. The qualitative description of each category was modified to reflect the SOLO taxonomy and the specific requirements of the assignments in particular the relation of theory to practice. Two sets of columns of qualitative

criteria were used. Criteria derived from Brown et al (1997) use categories that in some cases are as wide as 30%. The criteria developed by Johnson (1993) are given in bands of 10%. These were added where appropriate to assist markers in the process of finer grading within the wider bands. Guidelines on the process of marking were also provided (see box 1.8) to minimise halo and systematic order effects (Biggs, 1999; Brown et al, 1997).

Box 1.8: Marking process recommendations for external markers (from Biggs, 1999; Brown et al, 1997)

- Mark intensively until you have the criteria fixed in your head, then you can mark reliably a few questions at a time between other tasks
- At the beginning of each marking session (if there has been a gap since the previous session) re mark a few scripts
- Grade coarsely at first (qualitatively) by skim reading all the scripts and place in piles according to criterion categories. Then re-read with the criteria and mind to give quantitative value. Be prepared to change scripts at the borderlines of each category
- Shuffle the scripts between first and second readings
- Use the whole range of grades between 0 and 100%

The marking protocols for each assignment are different from each other in that each protocol has some subject specific examples of areas of knowledge. This level of difference does not reflect a difference in the level of understanding or type of knowledge assessed. The literature review marking protocol is more substantially different. The structure of the assignment is such that the functional knowledge is assessed separately from the declarative knowledge and the learning outcomes include uni and multi-structural outcomes. The written part of the assignment is therefore almost solely concerned with assessing declarative knowledge. Both the second and third assignments have only one part in which the emphasis is on assessing functional knowledge.

Framework category Skills, Personal and Propositional Knowledge – Groupwork assessment

The goal of practising empathy was considered as part of the goal of collaborating productively in groups. Problem Based Learning places great emphasis on group or teamwork. It is argued that the process of collaboration improves the effectiveness of learning and the effectiveness of

the individual in future collaborative settings (Myers Kelson & Distlehorst, 2000). The claim that Problem Based Learning improves group work skills and that this improvement produces measurable increases in learning and thinking and later on in patient care appears to be an assumption that requires further testing by research (Thomas, 1997). Given the importance attached to group work in the Problem Based Learning literature, there appears to be a deficit of summative studies of group performance in the Problem Based Learning literature.

An attempt was therefore made to assess this aspect of student performance using video assessment of each group undertaking a series of problem solving tasks. The studio facility used was based on one of the University sites. The video assessment was carried out on the last day of each groups programme. The groups were informed in advance that the exercise was being conducted. On the day each group was taken into the studio facility. The group sat in a semi-circle around a small desk. A flip chart and pens were made available. The audio-visual technicians provided a briefing on the technical aspects of the recording process and visual and sound checks were undertaken. The Principal Investigator gave a briefing and instructions to each group. Identical instructions were given on each occasion. The Principal Investigator watched the groups from the studio control room and interrupted groups only if they violated any of the rules laid down for each problem solving exercise. The video was recorded onto a master tape using one fixed and two roving cameras. The Principal Investigator and control technician selected shots from the live feed. The master tape was then edited onto a VHS tape showing each group performance in full.

The problem solving exercises were compiled from problem solving texts. The exercises were selected to provide a mixture of paper based and physical problems that were not directly related to the participants workplace. The problems also varied in the extent to which they required logical, practical and/or spacial awareness. It should be noted that the exercises were not designed specifically to test problem solving ability but rather to stimulate the group to use its collective skills/ knowledge/ abilities to solve the problems i.e. to perform as a group. The

exercises were not formulated as clinical 'scenarios' or triggers in order to minimise cueing advantage to Problem Based Learning groups.

The task of evaluating how well a team or group functions could be viewed as a simple task of measuring how effective a group is at achieving the objectives that it is set. However the real world is rarely as simple as this as groups are dynamic, tasks vary in complexity and groups work in different and complex contexts. The literature on group work assessment has therefore focused on identifying the kinds of activities/ characteristics/behaviours/ attitudes which individuals in groups and groups themselves need to develop to perform successfully in complex settings. Developments in measurement have proceeded alongside the identification of these characteristics.

Attempts were made to identify tools that could be used to measure how effective a group is at working together both in the Problem Based Learning literature and more widely in the literature on group work. Both in the Problem Based Learning and wider group work literature the majority of instruments identified were primarily for the use of group members themselves in the process of evaluating group performance for formative purposes. Examples include the 'Group Enrichment Task' (Woods, 1995), the 'Small Group Teaching Evaluation' used at McMaster University (Jaques, 1990), the 'Team Orientation and Behaviour Inventory' (Goodstein et al, 1983), and the 'The Interpersonal Perception Scale (IPS) (Patton et al, 1989). However, a small number of summative assessment instruments were identified.

The Tutotest is a standardised instrument designed to assess the skills and attitudes of medical students working in tutorials in a problem-based curriculum. The test was developed at Universite de Sherbrooke Faculty of Medicine (Canada) and has reported data on reliability and validity (Herbert & Bravo, 1996). The instrument consists of 44 items in four domains 'Effectiveness in Group', 'Communication and Leadership skills', 'Scientific Curiosity' and 'Respect for Colleagues'. However the test is designed to measure individual rather than group performance, requires approximately 24 hours of observation for a group of 6-8 students and requires multiple observations to achieve a reasonable level of reliability.

The Faculty of Medicine and Health Sciences, Newcastle University, Australia developed an observational assessment tool that is used both formatively and summatively to assess group process and group reasoning (Rolfe & Murphy L, 1994). The instrument is used to observe group performance during a specific group task and is carried out in two stages. The instrument consists of 22 criteria in three domains. No data is reported on reliability and or validity and contact with the authors confirms that no subsequent evaluation of the instrument has been carried out (*I Rolfe personal communication November 2001*). Each criterion is specified as a pair. The first behaviour is that which is considered appropriate, the second that which is considered inappropriate. The instrument also offers the possibility of assessing other outcomes of interest specifically 'Dealing with problems and making reasoned decisions in unfamiliar situations' and 'Reasoning critically and creatively'. The nature of the assessment task set in the video exercise meant that it would not have been possible to make judgements about all the criteria on the original instrument. Therefore a revised version was produced for this study (items in the instrument are given in box 3.7).

Two independent 'experts' carried out the assessment of the video footage using the instrument. One was a social scientist with experience of group observation techniques. The other was a professional training consultant whose training activity included providing training on team/group work. Neither had any experience of Problem Based Learning. The assessors were provided with an edited VHS video to analyze 'at home' independently of each other. Groups were identified on the video with a number. The assessors were therefore 'blind' to the allocation status (i.e. experiment or control) of each group.

With hindsight it seemed likely that exercises with multiple solutions and which may involve the making of value judgements were more likely to provoke behaviour that revealed a groups capabilities at working together. It was also unrealistic to require assessors to analyse more than 10 hours of video footage. It was therefore decided to focus the analysis only on the problem solving tasks that appeared to provoke the most discussion/ non- consensual debate amongst the

groups. The Principal Investigator reviewed all the video footage and three problems were identified in this category, 'The bomb scare', 'The line problem', and 'Build a bridge'. In the year two videos because the groups had been set a time limit for completion of all the exercises and these three problems were completed in approximately 15 minutes. They were therefore included on the assessors edited video in their entirety. No time limit was given to the first year groups and therefore they took longer to complete the exercises. In order to bring the length of video footage for these groups down to roughly the equivalent of the year two groups the video footage of these problems was edited to remove excess periods of silence or inactivity.

Framework category 'Approaches to Learning' - ASSIST

There are a variety of ways in which capabilities for self –directed learning can be conceptualised and measured. Two instruments were used in the study. The short version of ASSIST (Approaches and Study Skill Inventory for Students) will be discussed in this section. The Self Directed Learning Readiness Scale (SDLRS) was administered as part of the follow-up questionnaire and is discussed in the next section. ASSIST is based on a set of concepts that have emerged over a substantial period of time concerning student motivation, study methods, approaches to learning and learning strategies. ASSIST is a development of an earlier instrument the Approaches to Study Inventory (ASI). Both the ASI and ASSIST underwent extensive development and testing and have reported high levels of internal reliability across different student contexts. Relationships with academic performance are also fairly consistent with positive correlation normally found with the strategic approach and negative correlation's with both surface and apathetic approaches (Entwistle et al, 2000; Tait & Entwistle, 1996). The ASI has previously been used in studies of Problem Based Learning (Coles, 1985). The short form of ASSIST was used (18 items) as this was more likely to be acceptable to both students and teachers. There is evidence that this subset of 18 items provides adequate reliability for measuring deep and surface approaches to studying (Richardson , 1992). The conceptual basis of the instrument is that a deep strategic approach to learning is more likely to lead to conceptual

understanding. The 'deep' aspect relates specifically to the claim that Problem Based Learning helps students to develop 'richer mental models' whilst the 'strategic' aspect would seem to relate clearly to the goal of developing self-directed learning. The instrument was administered to participants at the beginning of the programme and again on completion of the programmes.

Longer term effects - Follow–up surveys

Consideration of the long term effects of any educational programme is an important aspect of measuring programme impact (Wilkes & Bligh, 1999). The question is whether improvement on some kind of assessment immediately on completion of the educational intervention actually translates subsequently into improved performance (Abrahamson, 1984). The issue is particularly pertinent where the educational programme has a direct vocational role i.e. the preparation and/or continuing development of practitioners in a particular field. It is quite possible that the impact of learning on practice may not become apparent to the learner (or the external observer) until some period after the conclusion of the educational programme (Pascarella & Terenzeni, 1991). Consequently follow-up studies may produce quite different results to those obtained at the immediate completion of the programme. Claims for the importance and /or legitimacy of Problem Based Learning usually emphasise the need to develop new kinds of practitioner, improve the performance of practitioners and/or improve student satisfaction (Albanese & Mitchell, 1993; Burns & Glenn S, 2000; Engel, 1991; Savin-Baden, 2000a; Savin-Baden, 2000b; Vernon D.T & Blake, 1993). The technical and methodological difficulties of assessing impact at this level of complexity coupled with the limited duration and funding of most educational evaluations means that there are comparatively few studies of this kind (Hutchinson, 1999).

The limited resources available to the project meant that the only possible method of data collection for the longer term follow-up was a postal survey. It is argued that six months is a period of time in which the quality of opinion about the utility of the programme is more likely to be experience based and less likely to be based on factors such as entertainment or prestige

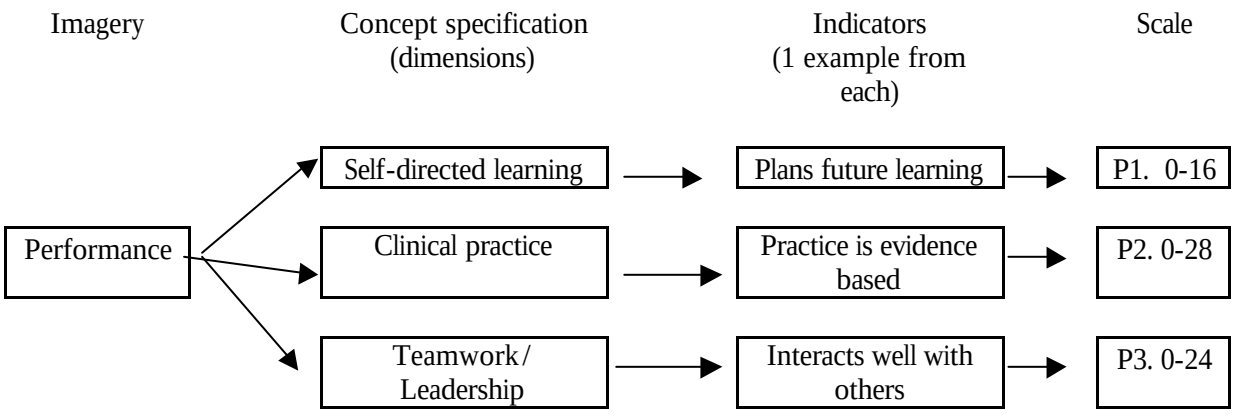
(Nowlen, 1988). In previous follow-up studies of Problem Based Learning that were identified the subjects were former undergraduate medical students' (Peters et al, 2000; Woodward & Ferrier, 1982b; Woodward et al, 1988). Therefore the instruments used in these studies were not applicable in this study. One of the most cited claims of evidence for the impact of Problem Based Learning on life long learning, used as an outcome measure doctors reports of their practice in managing a number of specific disease conditions (Shin et al, 1993). That Doctor's who had followed a Problem Based Learning programme in their medical training showed more up to date and evidence based knowledge and clinical management was used to argue that they must therefore have better self- directed/life long learning skills. Consideration was given to the use of such an approach in this study. Such an approach could not be used in this study for two reasons. Firstly, the participants worked in a variety of medical and surgical wards, which means that emphasis placed on particular aspects of clinical practice knowledge and management will vary. Secondly, there are very few issues in clinical nursing practice in which a defined path of clinical management is unambiguously agreed to be 'the best'.

In the absence of any appropriate pre-existing instruments, 'new' instruments were developed specifically for use in this study. The measurement instruments used were embedded in a questionnaire designed for use in a postal survey. Consideration was given to ease of and time for completion in order to minimise the likelihood of non-response and the return of incomplete questionnaires. A structured format that in the main used predetermined standardised response formats was selected to aid completion, increase reliability and facilitate data analysis. The questionnaire for former students' comprised of questions about any changes in their work role since completion of the programme, a set of statements designed to assess their performance, a set of statements designed to obtain their views about the impact of the programme on their practice, a set of statements designed to assess the readiness for self-directed learning and a set of statements designed to assess their views about the strengths and weaknesses of the programme.

Pilot student questionnaire

The initial operationalization of the concept of performance is illustrated in figure 1.3. The literature on Problem Based Learning makes claims about improved performance using various terminology that were distilled into the three broad dimensions used in the pilot student questionnaire; self directed learning, team work/ leadership, and clinical practice. It is recognised that performance in these areas is interlinked both conceptually and in practice and furthermore that assessing performance in areas such as these areas is highly problematic (Hutchinson, 1999; Van Der Vleuten, 1996).

Figure 1.3: Preliminary operationalisation of 'performance in practice' (after Bryman & Cramer, 1995)



A multi-item scale was created to assess performance in each dimension. Each scale used a number of items that were developed from tools used in previous studies on the impact of Problem Based Learning (Peters et al, 2000; Walton et al, 1997; Woodward & Ferrier, 1982a) and from other relevant performance assessment tools (Brown et al, 1997; Patton et al, 1989; Quinn et al, 1990; Redding, 1992). The items in each dimension are given in the tables 1.9 to 1.11 below.

The items in each dimension were constructed so that a strong response on one item (5 or 1) is consistent with a strong response to the opposite (1 or 5) on its pair item. The optimal positive performance rating for a dimension with 4 pairs of item is therefore 16. A combined score near to zero suggest a lack of consistency and/or indecision on the part of the respondent.

Table 1.9: Pilot student follow-up questionnaire: 'Self directed learning' performance items

Indicator 1	Indicator 2	Scoring
I have not learnt anything new since the last course I went on	I am proactive with regard to my learning and development	Indicator2 – Indicator1
I am not confident in my ability to interpret the scientific strength of clinical research evidence	Compared to other nurses in my ward I frequently use research articles	Indicator2- Indicator1
When being given an explanation or instructions I hesitate to ask questions, because I don't want to appear uninformed or ignorant.	I get told off for asking too many questions	Indicator2-Indicator1
I don't plan for my personal and career development	My personal learning and development is my responsibility	Indicator2-Indicator1

Scale: Optimal positive score per item = 4 – Optimum positive dimension score = 16.

Table 1.10: Pilot Student follow-up questionnaire: 'Clinical practice performance' items

Indicator1	Indicator 2	Scoring
I refer patients with complex problems to other members of the healthcare team as quickly as possible	I am confident in my ability to care for a patient with complex clinical problems	Indicator2-Indicator1
I am not confident in caring for patients from different ethnic or faith backgrounds to my own	I always assess and plan care to meet patients social and spiritual needs	Indicator2-Indicator1
I do not know much about national health or nursing policy	I keep my self abreast of developments in health policy and practice	Indicator2-Indicator1
Management of the patient's medical problem is my priority	Preventative health care is a very important part of my practice	Indicator2-factor1
Listening/talking to patients is a lower priority than managing their clinical care	I develop good relationships with the patients I care for	Indicator2-Indicator1
When I need advice/ information I turn to colleagues/ experts in my hospital and usually do what they say	My practice (knowledge) is evidence based	Indicator2-Indicator1

Scale: Optimal positive score per item = 4. Optimal positive dimension score = 24

Table 1.11: Pilot student follow-up questionnaire: 'Team member /leader performance' items

Sub dimension	Indicator	Indicator	Score
Teamworker	Indicator 3: I hold back from contributing in team meetings	Indicator1: I interact very well with other people	Indicators 1+2+4+6 – Indicators 3+5
	Indicator5 : In my position I get quicker action by doing a job myself rather than asking someone else to do it	Indicator 2: In making clinical decisions I like to consult with other members of the health care team frequently	
		Indicator 4: I am confident in my ability to work as a member of a team	
		Indicator6: I trust my colleagues/ team members	
Leadership	Indicator1 I withdraw when the other person confronts me about a controversial issue	Indicator2 I am confident in my ability to deal with conflict in my workplace	Indicators 2+4+6 – Indicators 1+3+5
	Indicator3: I expect others to participate without encouragement from me	Indicator4: I am concerned about what happens to other members of the team	
	Indicator5: I am not able to exert much influence over other people	Indicator6: People look to me for guidance and help	

Scale: Optimal positive score for dimension teamwork = 18. Optimal positive score for dimension leadership = 12

Pilot line managers questionnaire

As a form of triangulation student's immediate line managers were asked to rate their performance. The students varied with regard to their position in the organisational hierarchy, for example, some were ward managers and others junior staff nurses. This suggests that the person who has 'line managerial' responsibility for a particular participant will not always work with them sufficiently closely to be able to provide an assessment at the same level of detail as that required by the instruments in the student questionnaire. The multi-item assessment instrument used in 'the line manager' questionnaire was developed from other tools (Brown et al, 1997; Patton et al, 1989) used to assess performance of students in work related behaviours that the educational programme in the study claimed to develop. There are 12 items in the scale.

Each item attracts a score of between one and four. The possible scale for assessment of performance ranges between six and 48. A total score of above 30 indicates a positive (good) performance (this is the mid point between 12*2 and 12*3).

Pre-testing of pilot student follow-up questionnaire

Pre-testing of the questionnaire broadly followed the procedures outlined by the American Statistical Association (American Statistical Association, 1997). The paper outlining the development of the questionnaire and the questionnaire itself were made available from the project website and the project e-mail list used to ask for comments and feedback. The questionnaire was redrafted as the result of a small pilot study and the identification of further relevant literature.

The questionnaire was administered to a group of 13 students who were qualified nurses currently undertaking a programme to top-up their diploma to a BSc in mental health. The Principal Investigator administered the questionnaire in the classroom and conducted a debriefing session to identify any problems they had with completing the questionnaire. It took the students about 15 minutes to complete the questionnaire. Measures of internal consistency estimate how consistently individuals respond to the items within a scale. The internal consistency of the measurement scales within the questionnaire were assessed using SPSS to calculate Cronbach's Alphas (α) for each dimension and scale within the instrument. Negative scoring items were recoded for the purpose of analysis.

Cronbach's Alpha is the most commonly used estimate of internal consistency of items in a scale. The Alpha measures the extent to which item responses obtained at the same time correlate highly with each other. Alpha is a measure of level of mean intercorrelation weighted by variances, or a measure of mean intercorrelation for standardised data. The widely accepted

social science cut-off is that Alpha should be .70 or higher for a set of items to be considered consistent (Bryman & Cramer, 1995). An Alpha is calculated for the whole scale and for each item within the scale. An Alpha is also given which indicates what the overall scale Alpha would be if a particular item is removed. If the Alpha will be higher when an item is deleted, the researcher infers that the item is not tapping the same construct as all of the other items and therefore it should be removed from the scale (Garson, 2001).

Table 1.12: Student follow-up questionnaire pilot - Cronbach's ∝ coefficients

Dimension	Positive items coefficient	Negative items coefficient
Performance (Total)	0.68	0.7
Self directed learning	0.25	0.25
Teamwork	0.16	0.54
Leadership	0.38	0.63
Teamwork and leadership	0.49	0.6
Clinical Practice	0.43	-0.87

Table 1.12 gives the alpha for the positive and negative items in each scale separately Alphas for the total on each dimension were calculated by re-coding the negative items. For these 'total' items only the scale measuring 'teamwork and leadership approached a reasonable level of internal consistency (∝0.69). The scale measuring the dimension 'clinical practice' was negatively correlated with an ∝>-1. It is possible that given the small number of items and small sample that while the true population covariances among items are positive, sampling error produced a negative average covariance. Or it may simply be the case that the items do not truly have positive covariances, and therefore may not form a useful single scale because they are not measuring the same thing (Nichols, 2002).

Pre-testing of pilot line managers follow-up questionnaire

Pre-testing was carried out on a small sample of nurse managers (n=7) at a local NHS Acute Hospital Trust. The Principal Investigator gave out the questionnaire at a meeting and the managers completed them at the time. The questionnaire took approximately 10 minutes to

complete and no problems were reported. The Cronbach's Alpha for the scale in the questionnaire was 0.97, which suggests a satisfactory level of consistency. It is interesting to note that there was no statistically significant correlation between the overall score given for global assessment of performance and the mean item score. In 3 cases the global assessment and item mean were identical. In 2 cases the global assessment was higher and in 2 cases lower.

Development of 2nd version of questionnaires

The combination of internal and external review, the identification of other relevant literature and the results of pre-testing indicated that substantial modification to the student questionnaire was required. A systematic review of research evidence and best practice in questionnaire design became available in early 2002 (McColl et al, 2001). The revisions carried out on the basis of this report are summarised in box 1.9. On the sub-scales teamwork, leadership and clinical practice the removal of items with low Alphas scores and/or with possible confusing negative wording left 21 items remaining. These were revised into a single 21 item scale measuring the dimension 'Capability for Clinical Practice Organisation'.

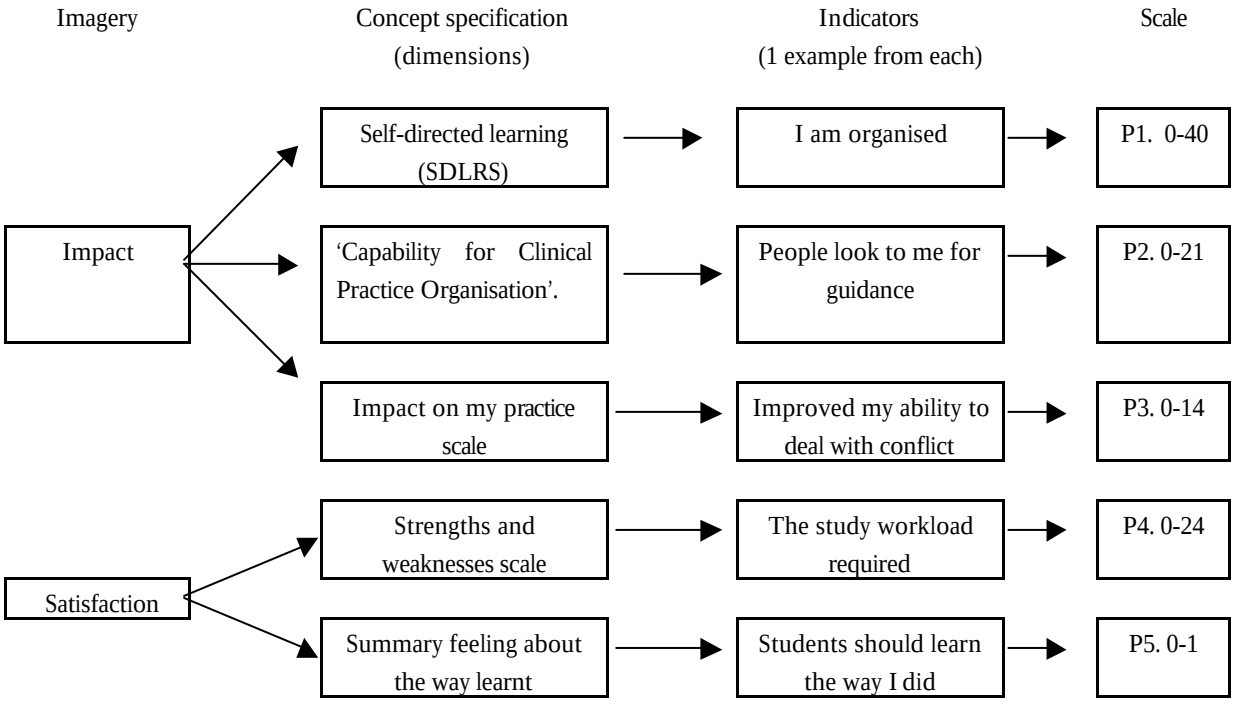
Box 1.9: Revisions to the follow-up questionnaires based on the systematic review of questionnaire design (McColl et al 2001)

- Questions about changes in the workplace and participation in practice development moved toward the end of the questionnaire (can be regarded as threatening)
- Removal or rewording of negatively worded items in the measurement scales
- attention to consistency in use of question stems within items and scale category headings
- Use of numbers instead of tick boxes on scale categories.

The Self Directed Learning Readiness Scale (SDLRS) (Fisher et al, 2001) measures the degree to which an individual possesses the attitudes, abilities and personality characteristics necessary for self directed learning. The instrument was developed by nurse educators in Australia using a rigorous three-stage process. In the first stage a bank of 93 items were developed from the existing literature. In the second stage a two round modified Delphi technique was used in which

selected experts independently identified those items that they felt were necessary for self-directed learning. In the third stage pretesting of the SDLRS, the final selection of items was carried out using item-total correlation based on data from a sample of 201 nursing students. Items with a corrected item-total correlation score of <0.3 were removed removed from the scale leaving a 40 item scale with an alpha for the total item scale of 0.924. Factor analysis identified three component subscales, Self Management (n= 13 items, ∝ 0.857), Desire for Learning (n= 12 items, ∝ 0.847), Self Control (n=15 items, ∝0.830). Based on the pilot study results the authors argue that a score of 150+ indicates a readiness for Self Directed Learning. The SDLRS instrument was included in the revised student questionnaire the operationalisation of which is illustrated in figure 1.4 on the following page.

Figure 1.4. Operationalisation of concepts in 2nd (final) student follow-up questionnaire



The questionnaire for managers/supervisors was also revised using the systematic review referred to above. The scale was remodelled to include additional items from the Clinical Supervisors report form developed to assess practice performance of medical students in the Problem Based Learning programme at the University of Newcastle (NSW) Medical school (Saunders et al, 1982).

Pre-testing of the 2nd version of the student follow-up questionnaire

The pre-testing of the second version of the questionnaire followed the process used in the pre-test of the first version. The subjects were 22 qualified nurses undertaking post registration nursing education programmes at Middlesex University. The first group had recently commenced a level 3 module 'working with people with dementing illnesses and their carers'. The second group was two- thirds of the way through a level 3 programme for advanced practitioners in Accident and Emergency nursing.

Table 1.13: Student follow-up questionnaire pretesting of final version - Cronbach's ∝ coefficients

Dimension / Scale	No. Cases	Cronbach's Alpha
'Capability for Clinical Practice Organisation'.	20	0.7518
Self Directed Learning Readiness Scale	22	0.9156
Impact on my practice Scale	21	0.8588
Programme strengths & weaknesses	21	0.8398

The results of the analysis of the internal consistency of the four different scales in the redesigned questionnaire are given in table 1.13. The consistency of all the scales appears satisfactory. Further analysis of the scale 'Capability for Clinical Practice Organisation' revealed that Alphas for the two groups were quite different with the Alpha for one group being 0.5399 and the other 0.8582. For this reason it was decided that the scale would be not be modified further. The total score for the 'Programme strengths and weaknesses' scale showed a statistically significant positive correlation with the students overall rating of how they learnt on the programme ($r=0.637$ – Significant at 0.01 on a 2 –tailed test), providing some evidence of the validity of the scale items.

Administration of the follow-up questionnaires

The questionnaires were sent to all students who completed the programmes and to the person who they named as their line manager at the time the questionnaire was sent. Each student was contacted prior to the questionnaire being sent to inform them that the questionnaire was being sent and to check that the contact details for them and their line manager were up to date. The questionnaires were sent by post with a personalised covering letter and a prepaid return envelope to maximise response rates. Where possible non respondents were contacted by telephone and additional reminder questionnaires were sent where required. The questionnaires were sent to the first cohort of students approximately 8 months after they completed their programme. The questionnaires were sent to the second group students approximately 4 months after they completed their programme.

Timing of data collection

The principles used for deciding the timing of data collection from students varied for each particular instrument. The timing of the data collection for each group was planned in advance as illustrated in table 1.14 below.

Table 1.14: data collection timetable module 1, 2001-2002

Group	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
A25SGL					2	3		1&4				
A25 PBL					2	3	4			1		
A24SGL					2	3	4		1			
A24PBL					2	3			4		1	

Key: 1= CSQ. 2 = ASSIST. 3= Student workload. 4= Observation

No data was collected from students in the first four sessions in each academic year. Sessions 11 and 12 of the first module were used for students presentations done as part of their assessment

so data was not collected in these sessions. The first ASSIST questionnaire was completed at the fifth session and the second at one of the last sessions in the academic year. The Course Evaluation Questionnaire was completed at the end of each module. The timing of the administration of the student workload questionnaire was decided randomly for each separate group. For data collection from students the instruments were distributed and collected by the Principal Investigator at an appropriate time during a teaching session. Where possible data was collected from all students members of each group at the same time. Where a student was absent they were asked to complete the instrument at the next session they attended

Economic evaluation

An important consideration in the evaluation of any teaching and learning strategy in a climate where the resources available for the creation of learning are scarce relative to the demands made upon them is the relative costs of any benefits obtained from using a particular strategy. The basic framework for economic appraisal is that all interventions require resources that have alternative uses and therefore involve a sacrifice of benefit elsewhere (cost). At the same time all effective interventions achieve results that are of value (benefits). The process of weighing gains against sacrifices is known as the cost-benefit approach (Drummond, 1980). Obviously different perspectives can be taken on what is a cost and what is a benefit. To a teacher a finding that students do more 'homework' may be viewed as a benefit, whilst to students themselves this may be viewed as a cost. The value of any cost benefit analysis is only as good as the data upon which such estimates are based. Data on a range of student outcomes that can be construed as 'benefits' were collected. Given that one of the major concerns expressed in the Problem Based Learning literature is that 'Problem Based Learning is more 'expensive' (Berkson, 1993), the 'cost' focus was a comparison of teacher 'workload' between the two curricula. All the teachers involved in the study contributed to the development of the experimental (PBL) curriculum. The control (SGL) curriculum was already in existence. Therefore the focus of the data collection was on teacher 'workload' associated with 'delivery' and support of students during 'term' time. The

tutors were provided with a form to record programme associated workload on a weekly basis. Initially teachers were emailed on a weekly basis to remind them to complete their forms. However, this proved counterproductive as it irritated the teachers. It is likely that the teachers did not complete these forms contemporaneously.

General Strategy for data analysis

The main purpose of the qualitative data collected in the study was to provide an account of the process of programme design and implementation. It will also contribute to a greater understanding of issues of student and teacher satisfaction and in the interpretation of the study results. The methods of data collection and analysis used for these purposes are discussed in Part II of the report that focuses on these issues. The strategy for the analysis of quantitative data follows the steps described by Altman (1991) and is summarized in box 1.10. The reporting of the study results conforms to the requirements of the CONSORT statement (Moher et al, 2001). Details of analysis issues specific to each instrument are reported along with the results of each particular outcome/ measure. Analysis was undertaken on the basis of the full number of participants who entered the trial. Sensitivity analysis involved analyzing the data with the 'worst' outcomes attributed to those participants that failed to complete the programme (Altman, 1991).

Box 1.10: Strategy for data analysis (from Altman, 1991)
1) Data collection
2) Data entry
3) Data checking
4) Data screening
5) Data analysis
6 Checking results
7 Interpretation

Data were checked for errors in either coding or data entry, using range checking, paired variable checks, checking for outliers, and inter-item consistency. Data were screened to establish and where necessary manipulate the distribution of the data for each variable. Data were assessed visually using 'Normal Plots' and statistically using the inter-quartile range, standard deviation and Shapiro Wilks W test for normality. Where data was not normally distributed and/or where the

data is likely to be biased by the effect of atypical values (outliers) log transformations were used and data analysis carried out on the transformed data.

The principle focus of data analysis was to summarise the completed observations to provide answers to the research questions (Selltiz et al, 1976). The principle aim of the study was to investigate the effects of Problem Based Learning on student 'attainment' (i.e. benefit) and on teacher's satisfaction and workload. Data analysis therefore focused on establishing whether there was variance in outcomes (Taylor Fitzgibbon, 1996). The analysis included both descriptive and comparative analysis. Description includes the use of visual presentation of the data in formats appropriate to the type of data and quantification of variability. For the purpose of data analysis the study design can be characterised as a mixed between methods design (Kinnear & Gray, 2000). The main analysis focused on comparison of the pre-specified outcome measures between the experimental (PBL) and control (SGL) groups. The approach to data analysis focused on estimating effect size rather than hypothesis testing with confidence intervals constructed for each result (Altman, 1991). Where results were not as expected i.e. as hypothesised, data were rechecked and re-screened to assess whether this variance was due to a true difference in the outcome that was being measured (Selltiz et al, 1976).

Part I of the report provided the rationale for the research, an overview of the principles of Problem Based Learning, the rationale and description of the research methodology. Part II describes and analyses the experimental (PBL) and control (SGL) curricula and the contexts in which the research was conducted. In so doing key ideas in the practice of Problem Based Learning are elucidated. The process of implementation student and teacher responses to Problem Based Learning are also analysed.

PART II

STUDY CONTEXTS, PARTICIPANTS, CURRICULUM THEORY AND PRACTICE

Part II Contents

	Page
Introduction	**70**
Sources of 'process' information / data	73
The Specific institutional and disciplinary contexts of the study	**76**
Teacher and student cultures and relationships in the study programmes	**80**
Guidance & support for learning	82
Teachers Previous knowledge and experience	84
Teacher beliefs, values and expectations	87
Student personal, professional and educational background	91
Student motivation and expectations	95
Differences in student characteristics between the experimental and control curricula	98
Student participation and morale	99
Programme content and assessment in the study	**102**
Programme aims and intended learning outcomes	102
Assessment and feedback in both curricula	106
Student workload and opportunities for practice in the experimental and control curricula	108
The control (SGL) curriculum in operation	**112**
Teaching methods and styles in the control (SGL) curriculum	112
Design and organisation of the experimental (PBL) curriculum	**115**
Preparation of students for Problem Based Learning	118
Creating the environment for the development of skills and expertise	119
Development of scenarios	123
Organisation of the small groups in the Experimental (PBL) curriculum	127
Role of the teacher in the Experimental (PBL) curriculum	130
Preparation of the teachers for the role of Problem Based Learning tutor	132
Teachers performance in the experimental (PBL) curriculum	133
Students response in the experimental (PBL) curriculum	134
Exploring differences between the two curricula -	**139**
Classroom interaction in the experimental and control curricula	139
Summary of similarities and differences between the Curricula	144

Figures

Figure 2.1 Concepts of the inner learning environment and specific Problem Based Learning dimensions	72
Figure 2.2 Example learning contract provided for students	82
Figure 2.3: Dichotomous model of subject/facilitator expertise	86
Figure 2.4 Subject/ Expertise balance amongst teachers at the beginning of the study	87
Figure 2.5: Activities undertaken during private study time by curriculum group	110
Figure2.6: Purpose of studying for assessment or not	111
Figure 2.7: Timetable for a module in the control (SGL) curriculum	114
Figure 2.8: Module plan from the experimental (PBL) Curriculum	121
Figure 2.9: Structures, actions and products of small group learning	128
Figure 2.10: Classroom Interaction: experimental (PBL) curriculum - Teacher Y 22/2/2001	141
Figure 2.11 Classroom Interaction: Experimental (PBL) Curriculum - Teacher X 21/3/2002	141
Figure 2.12: Classroom Interaction: Control (SGL) curriculum teacher 1 31/1/2002	142
Figure 2.13: Classroom Interaction: Control (SGL) curriculum teacher 2 20/3/2001	142
Figure 2.14 Classroom Interaction: Control (SGL) curriculum, teacher 3 16/4/2002	143
Figure 2.15 Comparing patterns of classroom interaction analysis in the PBL and SGL curricula	143

Page

Boxes

Box 2.1: University/ Departmental General aims of Post Registration Education Provision: 104
Box 2.2: Advanced diploma in Medical/Surgical Nursing programme aims and learning outcomes 105
Box 2.3 Generic Problem Based Learning Essentials (Barrows 2003) 115
Box 2.4: Woods' nine-stage model of Problem Based Learning programme development 116
Box 2.5: Experimental (PBL) curriculum framework 117
Box 2.6: The Seven step PBL process used at Maastricht University (Schmidt, 1983) 122
Box 2.7: The eight tasks of PBL (based on Wolff, 2000) 123
Box 2.8: questions used to facilitate scenario design 124
Box 2.9: example of scenario from the Problem Based Learning curriculum: Nurse Sue Downs 126
Box 2.10: List of scenarios in the experimental (PBL) curriculum 127
Box 2.11: Teaching techniques used by the PBL teacher (after Wolff, 2000) 131
Box 2.12: Flanders' Interaction Analysis Categories (FIAC) 140
Box 2.13: Similarities and differences between the experimental (PBL) and control (SGL) curricula 146

Tables

Table 2.1: Teachers qualifications and experience at the start of the study 85
Table 2.2: Teacher scores on MRIQ scales 88
Table 2.3: Characteristics of students in the study by curriculum group 93
Table2.4: Self reported student motivation for undertaking course 95
Table 2.5: Response rates to student workload questionnaires 109
Table2.6 Difference (*d*= effect size) mean amount of time spent in spent in 'private' study per week 110
Table 2.7: Group size at the start and finish of the programme 129
Table 2.8 Experimental (PBL) curriculum groups progress in the various properties of group development 137

Part II

Introduction

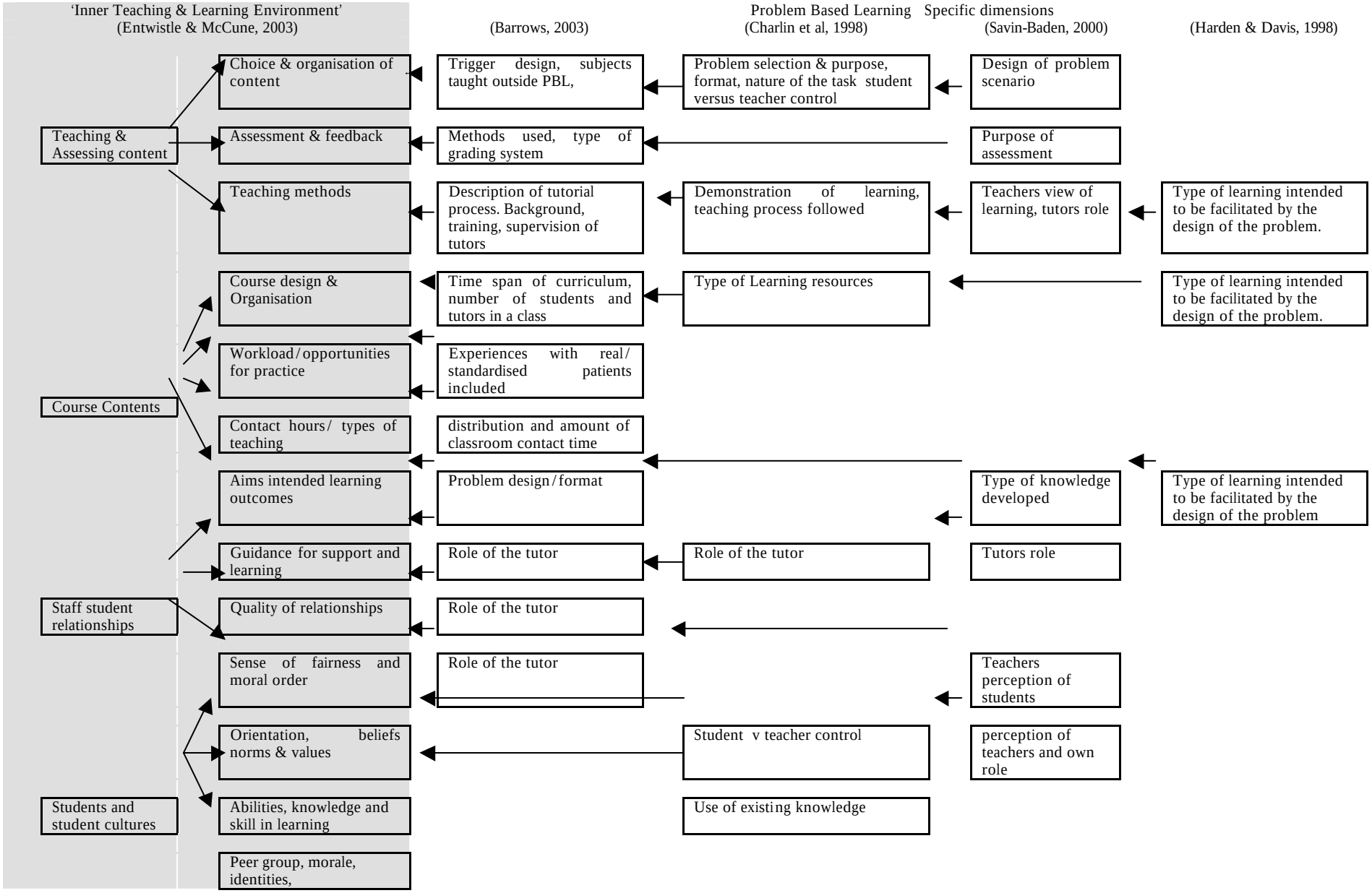
Evaluative studies in education can have a variety of purposes (Nevo, 1986). According to Cervero (1985), in the context of professional education they should seek to answer the complex question of under what conditions and for which types of individuals, which educational activities produce which changes in behaviour and improvement in client outcomes. The products of the study should contribute to the knowledge base for Teaching and Learning. Such a knowledge base would necessarily consider the conditions of learning, the properties of learning environments, the properties and characteristics of participants and the nature of their interactions (Desforges, 2001). The purpose of this part of the report is therefore to open up the metaphorical 'black box' of the contexts in which the study took place and the teaching and learning approaches used. And in so doing to analyse the differences the differences between the two curriculum approaches being compared.

Quite how the 'black box' should be conceptualised and its boundaries defined is open to differing interpretations. A wide variety of 'salient features' of teaching and learning are identified as important in the literature on Curriculum, Teaching and Learning in Higher Education generally and in the literature on Problem Based Learning. Whilst there are many areas of common ground there are also areas of dispute. It is not the intention here to try and justify a claim that the approaches used in either the experimental or control curricula were 'such and such' a teaching or curriculum approach (readers can decide for themselves). But rather to describe and analyse the approaches used in each curriculum across a sufficient number of characteristics to illustrate the links between Problem Based Learning and concepts of teaching and learning in higher education more broadly, and to articulate clearly the differences between the curricula. This is done both by analysing the curricula approaches using as a framework a general conception of 'the inner learning-teaching environment' in Higher Education and also as a specific form of Problem Based Learning

The model of the 'Inner teaching-learning environment in Higher Education' proposed by Entwistle and McCune (2003), was developed from theory and research on students learning. In this respect the model shares a similar conceptual basis to Problem Based Learning (see figure 1.1 in part I). Arguably, Problem Based Learning can be conceptualised as a systematic approach to manipulation of the 'inner teaching-learning environment'. Figure 2.1 shows four of the five key domains of the 'inner teaching and learning environment' and the key concepts in each domain (the shaded area). The fifth domain refers to the 'Institutional and Disciplinary Contexts' within which he other domains are enacted. The dimensions of Problem Based Learning highlighted by different authors have been mapped onto the Entwistle/McCune model. The model can be therefore be used to locate the specific example of Problem Based Learning used in this study within a wider Higher Education context. To do so the model is used as a framework for reporting the theory, practice and values of the two curricula being compared.

As figure 2.1 on the following page illustrates, different authors emphasise a number of different features of Problem Based Learning. The relative importance attributed to these features varies. It does not appear that the relative importance of each aspect has been demonstrated empirically in the context of Problem Based Learning. Given this variation, the characteristics of the experimental (PBL) curriculum is described as fully as possible in order to locate this particular example in the range of approaches that are called 'Problem Based Learning'. This section of the report is structured according to the component parts of the model, 'Institutional and Disciplinary Contexts', 'Staff and Student Relationships and cultures', 'Course and assessment contents', 'Course design and teaching Content.' The development and operation of the experimental (PBL) curriculum is discussed in detail and differences between the curricula explored. Any characterisations of the two curricula, the teachers who taught them and the students who took part in the study are structured and interpreted through the lens of the researcher. The characterisations attempt to provide simplified general trends of complex social phenomena, which are dynamic and interactional. Therefore the account is inevitably a partial a representation of a complex reality.

Figure 2.1 The 'Inner Teaching & Learning Environment' and its relation to specific Problem Based Learning dimensions



Sources of 'process' information / data

In complex educational interventions describing the 'process' is itself a part of the empirical study. The information on the programmes presented here was gathered from a variety of sources. Programme handbooks and learning materials, discussions with the teachers involved in the 'delivery' of the teaching both during the development of the curriculum materials and during the delivery of the programmes. Data on the students' demographic characteristics, professional qualifications and experience, and educational ability was collected by means of a questionnaire completed by all the study participants at the beginning of the programme. Combined with data from the field notes and classroom observations these provide insight into students expectations, motivations and abilities

Teachers' practices and expectations of students are framed within their own pedagogical repertoire of professional and personal knowledge and experience. Whilst these expectations may be tacit, they guide educators practice, relationships, and expectations of students (Millies, 1992). The experimental (PBL) and control (SGL) curricula operated (it transpired) within different pedagogical frameworks. However, the teacher's previous experience, values and expectations, both clinical and pedagogical influenced the interpretation of these frameworks and thus the teacher's practice. Knowledge of these characteristics, values and beliefs is therefore an important aid to interpreting the study results.

Data on the demographic and professional characteristics of the teachers was collected by a self-completed semi-structured questionnaire. The teachers also completed a questionnaire designed to assess teacher's beliefs about their role as a teacher. The 'My Role Is Questionnaire' (MRIQ) (Woods 1995) contains 18 items. Each item is composed of two or three contrasting statements designed to explore beliefs and attitudes that underlie teaching. The teacher is asked to allocate five marks between each statement that reflects the strength of their belief/ opinion

on each particular item. Each item is based on one or more specific pieces of research evidence about teaching and learning. Analysis involves the computation of scores on six scales.

An important aspect of establishing the validity of the evaluation is to monitor whether what is espoused in 'theory' e.g. in programme documentation is reflected in what actually happens in the classroom. Data collection during the programme included non-participant observation of randomly selected teaching sessions, field notes taken during meetings and discussions with teachers, and diaries kept by the teachers. Five observations were carried out for each group in each curriculum making 20 observations in each of the curricula. The tutors were informed a few days in advance that a particular session was to be observed. The researcher made notes during the teaching session and wrote these up immediately after the session was complete.

In the case of the classroom observations the researcher was not an active participant in the teaching and learning session. However, from the point of view of both the students but more especially the teacher the researcher was not perceived as a neutral observer, both teachers and students commented that they felt uncomfortable with the observers' presence. Teachers in the experimental (PBL) curriculum in particular were of the opinion that the presence of the observer affected the student's behaviour. Similarly in the field notes made during the course the project the nature of the interaction between the observer and the teachers takes place in the context of the situation where the observer is a colleague, leading the project and intimately involved in the development of the curriculum. The reflexive character of the qualitative fieldwork process means that the researcher inevitably injects something of themselves into the research process (Blaikie, 2000).

Qualitative data were analysed using the 'Framework Method' as described by Ritchie & Spencer (1994). This method is particular appropriate for identifying themes where there is an apriori set of issues identified. The characteristics of the 'inner teaching-learning environment'

identified in the Entwistle/McCune model were used as the framework (i.e. the apriori set of issues) for the analysis. The method consists of a four stage iterative process. In the first stage, the various data sources were read in a process of familiarisation. During this stage, notes were made which are the basis for the second stage of identifying thematic frameworks. The transcripts and notes are used to identify themes in response to pre-identified issues, themes that are introduced by the respondents (emergent issues) and themes that arise from recurrence or patterning in the data (analytic issues). Stages three and four involve the development of subheadings and headings and the mapping and interpretation of data from the separate cases under each heading. Stages two, three and four overlap as the process of analysis involves all three sets of activities simultaneously. In the reporting below quotations from the data are included that are illustrative of a particular theme/ issue.

The specific institutional and disciplinary contexts of the study

The specific institutional and disciplinary contexts in which the programmes in this study operate are that of English post registration nursing education. Therefore the relevant institutional contexts include The University, the National Health Service (NHS) and the Nursing Profession and specifically the interaction between the three. The University in which the study was completed is a former polytechnic and in English Higher Education policy jargon is therefore known as a 'post 1992' University. It is perhaps unwise to generalise about what this means in terms of teaching-learning environments. Institutions in the 'new university' sector tend to focus more on teaching than research. They also tend toward providing professional, vocationally oriented and 'newer' higher education subjects/disciplines. The University in this study includes traditional higher education subjects such as Sociology and History but predominantly provides more applied programmes. Examples of which include Theatre Design and Environmental Health.

'Nursing education' provision is somewhat divorced from the structural and cultural differences (real or imagined) associated with the divisions in English Higher Education highlighted above. In the UK nursing education is largely provided through University departments or colleges according to a nationally agreed frameworks of standards. But this is a fairly recent development. The move from Schools of Nursing attached to specific hospitals or health authorities coincided with the introduction in 1989 of the Diploma of Nursing & Midwifery in Higher Education (Burns & Glenn S, 2000). Many nurse teachers including those involved in teaching on the programmes in this study were originally nurse teachers in the different schools of nursing that through a process of mergers eventually formed the University department. There was a tendency amongst the teachers in this study to identify more readily with Health Service and Nursing rather than 'University' contexts.

The influence of disciplinary structures and cultures both in nursing and in the National Health Service (NHS) means that whilst the institutional culture shares many commonalities with the new university contexts there are some distinctive features specific to the discipline. Since the late 1990s education for health professionals (excluding medical education) has been managed on a quasi-contractual basis with firstly NHS Consortia and more recently NHS Workforce Confederations' being given the role of 'purchasers' of educational 'services' from Universities and other suppliers. This quasi-market arrangement operates through a quasi- contracting mechanism whereby the University provides individual NHS Trusts (hospitals or community services) with a 'menu' of programmes. An individual NHS Trust, through its local NHS Workforce Confederation then decides how many places it wants to purchase on a particular programme in the next academic year.

Although non- clinical health service and university administrators are involved in the contracting process, professionals, in this case nurses, are key actors on both the 'purchasing' and 'selling' side of this contractual process. The introduction of the contracting process could be viewed as formalising the division between 'education' and 'practice' in nursing (nurse managers tend to view themselves as 'nurses' rather than 'managers'). Whilst nursing cultures are dynamic, variable and complex (Suominen et al, 1997), cultures of practice tend to place a high value on 'doing' (Newman et al, 1998). This can be seen as part of a wider emphasis that nursing is primarily a practical rather than cognitive activity (Clarke, 1997). Thus whilst theories, models, research and evidence may seem important to nurse educators to the practitioner they often appear irrelevant to their task (Mulhall, 1997).

The clinical professions themselves are hierarchical, as are the organisations in which clinicians work. This is reflected in differences in status, power and salary (Dawson, 1994). In the English National Health Service, nursing roles are graded according to their responsibility. A nurse who has recently qualified would be appointed as a 'D' Grade staff Nurse. After a period in

practice, the nurse could apply for promotion to an 'E' grade senior staff nurse post either in the same ward or in a different ward or hospital. The 'F' grade post is considered a junior management post and nurse will often have the title 'Junior Sister'. The most senior ward based clinical post for a nurse is the ward manager (still sometimes called 'Sister' or 'Charge Nurse') and these posts are usually a 'G' grade. Levels of pay are determined both by grade and length of time in service. In a general surgical or medical ward the nursing staff will usually comprise of 1 'G' grade, 2 'F' Grades a larger number of 'E' grades and a larger number of 'D' grades who will be supported by a group of health care assistants. Theoretically the competencies required and the role of nurses in each grade should be similar from ward to ward and hospital to hospital However, it is clear that the local employment environment as well as an individual nurses experience and ability influences the grade at which as nurse is appointed. It should not be assumed therefore that the development of skills and competencies and level of responsibility either progresses smoothly and incrementally or that 'D', 'E', 'F', 'G' grade can be used as a simple proxy for intellectual or practical ability.

The relationship between the specific institutional contexts outlined above, student and teacher values, norms and action is complex. But it seems likely that they play some part in shaping the perceptions and values of students and teachers involved the study as they are part of these communities of practice (Lave & Wenger, 1991) and people adapt to organisational settings by being socialised into the prevalent norms (Eraut, 1994). Some of these contextual factors had quite explicit manifestations during the programme. For example, the institutional context seemed to shape teacher's negative perceptions of the likely reaction to the introduction of Problem Based Learning. The first cohort of students in one of the experimental (PBL) groups expressed their dissatisfaction early on. At a meeting one of the teachers was very concerned about this and asked the question

"What happens if they tell the managers?, the courses are being reviewed next year"

(field notes 11/10/00)

This concern seemed to have some justification in practice. Some students who were unhappy with Problem Based Learning did in fact complain to their workplace line manager (a ward manager). The ward manager took the complaint to a more senior manager in the National Health Service. In one case the manager contacted the Principal Investigator of the study directly. In the two other cases the National Health Service manager contacted a University manager who then contacted the Principal Investigator asking for a response to the complaints. In at least one of the instances the response of the managers on both the University and National Health Service side was to suggest that the matter should be referred to the most senior managers in the University department and the particular National Health Service Trust. In another example one of the teachers communicated to the researcher that

> "*she (a nurse manager) continues to intimate that the (hospital) is seriously considering withdrawing from the course and are using the research project as an excuse* "
>
> (Field notes 10/12/2001)

It would appear from this analysis that the wider institutional and disciplinary contexts may have influenced the outcomes of the study. The implications of this are discussed further in the discussion of the results in part III of the report.

Teacher and student cultures and relationships in the study programmes

The handbook produced for all students undertaking the programme (programme handbook) does not specifically discuss the issue of the relationship between teacher and student. Other than to emphasise that the students are practitioners and therefore must take responsibility for their own learning. The teachers and students signalled their views about the appropriate relationship by their comments and actions during the programme. There were similarities and differences between the teachers in their views and actions but generally the relationship between the teachers and students was discussed and operationalised as a form of mutually reciprocal 'ownership' by the teachers and students alike.

A teacher 'owned' a particular group of students in the sense that she thought of them as 'belonging' to her and whereas other groups 'belonged' to other teachers. Students felt that (or wanted to feel) their group 'belonged' to a particular teacher. In one of the first year Problem Based Learning groups the teacher made quite a lot of references to 'her' students both in discussion with colleagues and in discussions with students she was 'teaching' in the Problem Based Learning group. However, she was not talking about the students in 'her' Problem Based Learning group but students in another programme for which she was the programme leader. The students in the Problem Based Learning group later cited this as an example of why they perceived the teacher 'did not care' about their group.

In two out of the four groups in the experimental (PBL) curriculum, the student's frustration and anger about Problem Based Learning was turned on the teacher. For a period the relationships between the students and teacher in those groups appeared conflictual in nature and a detectable atmosphere of 'us and them' developed

> *" The students appeared to be deliberately excluding the teacher from the discussion by physically forming a circle of chairs that did not include her"*
>
> (PBL group teaching observation 14/11/00)

In this case the teacher recognised that her relationship with the group was not working and was visibly upset, reporting in a discussion with the study teaching team that she really wanted to 'chuck it all in'. In another Problem Based Learning group after another teacher had spoken quite firmly to a group member about his 'failure' to do the work he was supposed to do, the group went and did the work the student was supposed to have done. Then in the feedback session the group told the teacher that she was 'out of order' for speaking to him that in that way. The 'crisis' points in these relationships eventually passed. In all cases this coincided with one or more students leaving the group.

In the control (SGL) curriculum the separation between the teacher and the students was emphasised by the spatial division of the classroom and the teachers appropriation of the tools for teaching. In the experimental (PBL) curriculum this territorial separation was not evident. Everyone sat in a circle and it was usually the students who used the classroom equipment. Social interaction between the teachers and students in all groups was largely limited to the classroom. For example, with few exceptions teachers and students did not go to coffee or lunch breaks together. One of the control (SGL) curriculum groups the teacher did meet up with students in her group for social events for example at Christmas. However, in this case this action should be viewed in the context of a situation in which the teacher herself recognised that she felt some need to gain their approval

> *"I am filled with a deep rooted sense of needing to please students all the time,"*
>
> (Field notes SGL Teacher 1)

When rescheduling group meetings both teachers and students in the experimental (PBL) curriculum required that all the students, but not necessarily the teacher, should be able to attend

a rearranged date i.e. a ‘substitute’ teacher could be found. This can be seen as further indication of the separation between the 'the teacher' and 'the group'.

Guidance & support for learning

Guidance and support for learning can encompass any aspect of the educational environment. The University provides a range of general learning support services such as the library for example. University libraries are available on the site where students attended classes and in each of the hospitals where they worked. These ‘general’ support services were available to all students in the programmes. The experimental (PBL) curriculum was specifically designed to support learning and is discussed in detail below. The programme handbook lists a number of ‘methods’ that ‘will be used to facilitate student learning’ in the control (SGL) curriculum. Most are methods of 'teaching' that are discussed in more detail below. One ‘learning support’ activity listed is that the student will be required to draw –up a learning contract with the tutor and their clinical mentor at the beginning of each module. The value of learning contracts (Boud, 1991) was also recognised in the experimental (PBL) curriculum and the same criteria applied. However it became apparent that neither students nor teachers made use formal use of learning contracts in practice in either curriculum. In the second year of the experimental (PBL) curriculum students were actually given a proforma learning contract (see figure 2.2) below and time built into the ‘Problem Based Learning preparation days’ to allow them to complete and get comments on a set of learning objectives. However this too proved ineffective and most students did not formally develop personal learning objectives during the programme in either group.

Fig 2.2 Example learning contract provided for students

Learning Outcome	Strategy	Criteria For Assessment
Example: *Improve my confidence to speak out in meetings*	*Example* *Make a contribution every week in group discussions*	*Example* *Group members give positive feedback about my contribution*

Another specific means of support for learning available to students was the personal tutorial. The teachers offered students individual tutorial support on either an appointment basis or informally at then end of teaching sessions. However, the teachers frequently commented that the students did not take up the offer.

'few students take up the offer of supervision, which is not in keeping with the level of need'

(SGL Teacher 1 – field notes 13/1/2002)

Evidence to support this is found in the teacher's reports of the amount of time spent on personal tutorials. When expressed as amount of time per student who completed the programme the experimental (PBL) curriculum average was approximately two hours 21 minutes per student and the control (SGL) curriculum average approximately one hour 17 minutes per student for the whole programme. Ultimately the teachers in the control (SGL) curriculum scheduled individual tutorials for each student instead of 'teaching'. This did not happen in the experimental (PBL) curriculum perhaps because the students in the experimental (PBL) curriculum were more likely to seek tutorials in part at least because of their anxiety that they were 'missing out on something'.

In their evaluations students often made comment about the supportiveness or otherwise of their tutor. The majority of student evaluations in the control (SGL) curriculum referred to the supportiveness and encouragement of the teachers throughout their programme. By contrast students in the experimental (PBL) curriculum made frequent complaints about what they perceived as a lack of guidance and support from the teachers particularly during the early stages of their programme as they struggled to come to terms with the Problem Based Learning approach. On occasion (see above) these complaints deteriorated to the point where students accused the teachers of either being ignorant themselves or deliberately withholding information from them about what to do. However, the students who completed the experimental (PBL) curriculum appeared to recognise the support that the teachers had given

them

'the tutor was helpful in getting us through'
(Student evaluation form PBL group 2 yr. 2)

'…….(the tutor*) was fantastic, she made the PBL group a lot of fun*'
(Student follow-up questionnaire PBL group 1 year 2)

Teachers' previous knowledge and experience

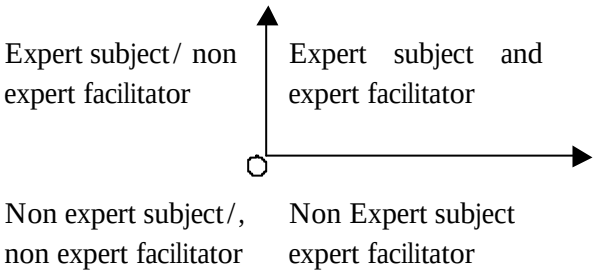
Five teachers taught on the programme. Information about their clinical and teaching qualifications and experience is given in table 2.1. All of the teachers who volunteered to take part were qualified nurses with extensive clinical experience. The control (SGL) curriculum was taught by the two 'Programme Leaders', who had volunteered their particular programme for the study. One left the University after the first year of the study and her teaching was taken over by a new member of staff who was newly recruited to the University as a Lecturer-Practitioner in Surgical Nursing. The experimental (PBL) curriculum was taught by two volunteer teachers who prior to their involvement had not been directly involved in the delivery of the programme. These teachers taught the experimental groups in both years of the study. The four teachers initially involved were all qualified teachers with extensive teaching experience including curriculum development. The Lecturer-Practitioner was not a qualified teacher and was undertaking her first teaching post. With the exception of the Lecturer-Practitioner, whose post was half – time teaching and half time clinical practice, all the teachers worked full time. None of the teachers described themselves as having been involved in research on teaching and learning although all had been involved in other types of academic and clinical research.

Table 2.1: Teachers qualifications and experience at the start of the study

Characteristic	Experimental group (PBL)	Control Group (PBL)
No. of teachers	2	3
Ages	56 47	53 40 38
Qualifications:		
i) Academic	1 degree, 1 Phd	1 degree, 2 Masters
ii) Clinical	Qualified nurses with post registration training	Qualified nurses with post registration training
iii) Teaching	Both qualified teachers	2 Qualified teachers
Teaching experience at start		
i)No. years as a teacher	15 16	17 8 0
ii)Months in current post	156 60	108 96 0
Clinical experience		
i)Years in clinical practice	4 20	13 11 13
ii)Clinical Specialty	Mixed	1 Mixed, 2 surgical only

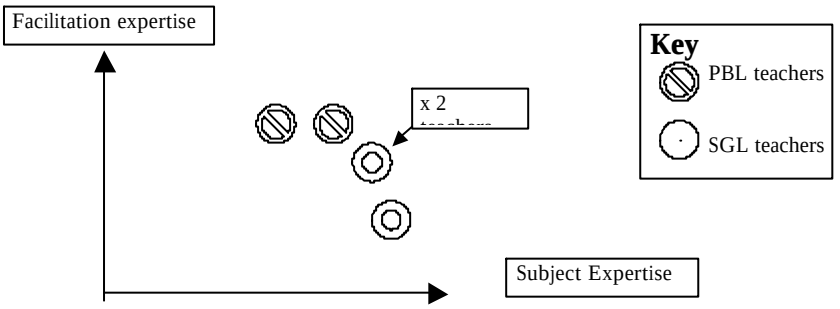
In the Problem Based Learning literature the issue of the tutor ability is usually discussed in terms of the tutors expertise in the substantive subject area and/or expertise in 'facilitation'. Much discussion has been generated on the issue of whether it is 'better' to have experts or non – subject experts as tutors (Silver & Wilkerson, 1991). The role of the Problem Based Learning tutor in theory and practice will be discussed in more detail below; here we are concerned with trying to locate the teachers involved in the project in terms of their expertise in the subject matter and in facilitation. The model implicit in discussions of the issue in the Problem Based Learning literature would appear to be something like the dichotomous model outlined in figure 2.3 below in which a tutor is either an expert or not. This maybe a useful framework for thinking about the issue in disciplines where students from one discipline are 'taught' by subject specialists from another. For example in traditional Medical Education programmes where medical students are traditionally 'taught' biochemistry by biochemists. It may also be useful in situations where students are 'taught' by academic subject experts rather than teachers who themselves are or have been practitioners.

Figure 2.3: Dichotomous model of subject/facilitator expertise



In the context of this study all except one of the teachers were qualified teachers and all were clinical practitioners in the same discipline as the students (Nursing). Whilst none of the teachers would necessarily claim that this makes them 'experts' either in the subject area of these programmes (i.e. 'specialist' medical/surgical nursing) or at facilitation it would suggest that at the very least they could not be regarded as having no expertise in the area. They all for example would be asked to teach as 'experts' by other teachers of post-registration programmes in specific subjects for example 'Wound Care'. The Principal Investigators' perception of the balance between the teacher's 'expertise' is presented visually in figure 2.4 below. The figure indicates that generally the Problem Based Learning teachers perceived themselves to be less expert in the specific subject areas covered by the course. With the exception of the 'new' teacher who joined the SGL curriculum, all the teachers regarded themselves as 'using' facilitation/ tutorial techniques in their teaching; in fact they all said that they used 'Problem Based Learning'. During the course of the study it became apparent that the teachers in the Problem Based Learning group did have specific subjects in which they possessed relevant 'expertise' and similarly, differential levels of expertise in facilitation/ tutorial skills became apparent. The Problem Based Learning tutors expertise as tutors/facilitators also increased as the study progressed.

Figure 2.4 Subject / Expertise balance amongst teachers at the beginning of the study



Teacher beliefs, values and expectations

The teachers beliefs, values and expectations about teaching and students were formally assessed using the 'My Role is Questionnaire' (Woods, 1995) and through their comments and actions during the project. The 'My Role is Questionnaire' has six scales each one of which is reported to assess a different aspect of teachers beliefs and expectations about their role as teachers. The scale labels used here are different to the labels used by Woods. The labels here provide a better indication of what each individual scale is attempting to measure. The 'Students as people' scale assesses the extent to which the teacher cares about the students as people. The 'Realising Potential' scale assesses the extent to which a teacher views her role as upholding 'standards' or as helping students' achieve their own potential. The 'Long Term Success' scale assesses the extent to which teachers care about students' long term success. The 'Focus on Learning or Teaching' scale assesses the extent to which the teacher focuses on student learning rather than teacher 'teaching'. The 'Empowering Students' scale assesses the extent to which a teacher is willing to empower students with part of the learning process. The 'Student Involvement in Goal Setting' scale assesses the extent to which the teacher is willing to allow students to participate in the process of setting goals and assessment criteria.

The figures in brackets after the scale name in the first column of table 2.2 are the average and the scale maximum scores supplied by Don Woods (*Don Woods personal communication 2000*). For

each scale a high score is preferable. None of the institutions from which the data on average scores is calculated is in the UK and the subject discipline areas were not given. Each scale is calculated by adding and subtracting the scores given by the teacher for particular items. The scores are presented separately for each teacher.

Table 2.2: Teacher scores on MRIQ scales (Individual, Mean (S.Dev)

Scale (Woods reported average – maximum scores)	Experimental (PBL curriculum teachers (n=2)		Control (SGL) curriculum teachers (n=3)		
	Teacher X	Teacher Y	i	ii	iii
(A) Students as people (22.5 – 30)	24	14	8	14	18
(B) Realising potential (7.7 - 10)	6	0	4	6	2
(C) Long term success (11.4 - 15)	13	15	11	9	3
(D)Focus on learning or teaching (29.1 – 40)	24	18	14	12	8
(E) Empowering students (11.4 - 30)	0	8	-2	0	2
(F) Student involvement in goal setting (3.5 – 5)	5	-1	3	5	1

The results on the scale (C) suggest that the teachers in the control (SGL) curriculum focused more on the success of their students in the context of the current course. This is perhaps not surprising given that they are 'Programme Leaders' who are ultimately accountable to The University authorities for the success of the programmes. With the exception of Problem Based Learning teacher X all the teachers had lower than average scores on scale (A) suggesting that they felt their primary responsibility lay with their students academic welfare and their responsibility stopped at the classroom door. With regard to their view of their role in terms of 'realising potential', Problem Based Learning teacher Y was at the extreme end of the continuum the result suggesting that she thought her role was primarily about upholding standards rather than realising individual potential. The views of teachers in the control (SGL) curriculum and Problem Based Learning teacher X were somewhere between the two. This impression is confirmed in scale (F) where Problem Based Learning teacher Y views goal and standard setting as a teacher rather than student responsibility and Problem Based Learning teacher X takes the opposite view.

The programme makes frequent reference to the fact that as qualified nurses the student's are expected to take responsibility for their own learning. Similar emphasise is given to the students' responsibility to reflect critically on their own experience and 'learning tools' are supplied for this purpose (learning contracts, reflective diaries). Teachers' frequent comments about what they perceived as students failure to take this responsibility during the course of the programme suggested that this expectation was 'real'.

> *"I feel frustrated by the students immaturity"*
> (Diary SGL Teacher 2 group yr. 1)

> *"The Teachers take responsibility to be on time, set up the classroom, do the preparation etc. and I get really frustrated with lack of commitment from students"*
> (Field notes PBL teacher X year 2)

However, the classroom practice of teachers in the control (SGL) curriculum consisted largely of teacher instruction (discussed in more detail below) and they often provided students with detailed written handouts suggesting that this desire was an aspiration rather than expectation on the part of the teachers.

The scores on scales (D) and (E) appear to be contradictory as one would expect a greater focus on learning as measured in scale (D) to be reflected in a greater orientations to empowering students with the learning process as measured in scale (E). Analysis of the measures of internal consistency estimate how consistently individuals respond to the items within a scale. The internal consistency of the measurement of these scales were assessed using SPSS to calculate Cronbach's Alphas ($\propto$) for each dimension and scale within the instrument. Cronbach's Alpha is the most commonly used estimate of internal consistency of items in a scale. The Alpha measures the extent to which item responses obtained at the same time correlate highly with each other. The widely accepted social science cut-off is that alpha should be .70 or higher for a set of items to be considered consistent (Bryman & Cramer, 1995). The Alpha scores for scales (D) and (E) are either negative or very low (scale D $\propto$ = -1.01 & -0.49,

scale E ∝ = 0.1 & -.83) suggesting these scales are not tapping the same construct (Garson, 2001). Thus it would be unwise to draw any particular inference from the scores on these scales.

The programme handbooks guide students as to what to expect from the programmes both in terms of process and content. The content of the programme as indicated by the handbook is discussed further in the section on 'Teaching and Assessing contexts' below. The handbooks discuss 'Learning' in some detail and activities that will be undertaken during the course are described as 'learning strategies' suggesting an orientation toward student learning rather than teacher 'teaching'. During the design of the experimental (PBL) curriculum there were many debates between the teachers about what an 'expert' medical/surgical nurse 'must' know and what the 'content' of the curriculum is or should be. This would suggest that in their 'usual' practice teachers followed a foundationalist approach giving evidence to 'covering to the subject' (Margetson, 1993).

This impression was confirmed during the teaching observations. In the control (SGL) curriculum the predominant classroom activity consisted of the teacher telling the students things, putting an emphasis on 'coverage' and thus teacher 'teaching'. The experimental (PBL) curriculum largely prevented the teachers from adopting the approach of 'telling'. However the Problem Based Learning tutors were very anxious about 'getting Problem Based Learning right' by which they seemed to mean them as tutors following the 'correct' procedure again suggesting a concern with teacher 'teaching'.

> *"I am very anxious about whether I am doing the right thing"*
> (Field notes PBL teacher Y. 11/2000)

In addition the teachers in the experimental (PBL) curriculum were concerned about their credibility with the students. They perceived their credibility was undermined because they could not 'tell' students what they 'knew'. Not surprisingly their concern about this appeared to diminished as the project progressed.

On the basis of the data presented above it would seem reasonable to characterise the model of teaching practice held by the teachers as the 'transfer' of knowledge and the 'shaping' of students to a predetermined pattern (Fox 1983). The teachers in who volunteered to 'teach' in the experimental (PBL) curriculum may have moved away from this model to some degree but the analysis suggests that that the differences between them and the teachers in the control (SGL) curriculum were not great. It is important to note that the experimental (PBL) curriculum, in theory at least, posed constraints on how the teachers could enact their preferred style of teaching.

Student personal, professional and educational background

Student entry characteristics are one of a number of factors affecting their eventual attainment (Mckeachie et al, 1986). In school age education is there empirical evidence to suggest that prior educational achievement will explain between 35-40% of variation in outcomes (Taylor Fitzgibbon, 1996). However, it is by no means clear that this relationship persists into adulthood or post-compulsory or professional education. It has also been argued that there are differences in the way that men and women value and perceive educational experiences. For example female students appear to value more highly the notion of 'a learning relationship', whereas male students tend to place more value on the teacher as a source of 'facts' (Case & Thomson, 1995), a difference that may be influential in student evaluations of Problem Based Learning curricula.

Data on participant characteristics were collected at the beginning of the academic programme using a self-completed questionnaire designed for this purpose. The questions were selected to provide information relevant to both establishing the external validity of the study and on

factors that are suggested in the educational literature to affect educational attainment. The questionnaire contained a mixture of closed and semi-structured questions. The response rate was 100% (n=67). Results from the questionnaire are given in table 2.3 below.

All participants in the study were qualified registered nurses (RN). The minimum professional education they will have undertaken is a three-year training programme that involved both practical and classroom based learning. The curriculum content and mode of delivery of this initial nursing programme has changed over time as ideas about the role of the nurse and nursing education have changed. Perhaps the most significant, which coincided with the relocation of nursing education from the NHS to Higher Education Institutions, was the development of the Diploma in Higher Education in Nursing. Prior to this Registered General Nurse (RGN) Training programmes were based on apprentice type models. The diploma course placed more emphasis health promotion, psychology, sociology and health policy and reduced the amount of time students spent on acquiring psychomotor skills in the University (Burns & Glenn S, 2000). However, within a general framework Diploma programmes vary in their execution according to a range of factors including the interests and orientations of the programme staff and the students.

The minimum educational entrance requirement to undertake the RGN or Diploma programmes is five O' levels or GCSE's (school exams taken at age 15/16) or by passing the General Nursing Council entrance test. Some nurses will have obtained the RN qualification by taking a 'conversion course' from 'Enrolled Nurse' (a qualification that no longer exists). After qualification nurses can take a variety of educational programmes from one day training courses to degree level short programmes that are accredited by a national board (Formerly the English National Board for Nursing Midwifery and Health Visiting). The degree is becoming the benchmark currency for promotion to higher grades in nursing and also has more currency outside of nursing. Much post-registration education (including the programmes in this study)

have become organised within a framework that allows students to use the modules on the course towards 'topping –up' their qualification from diploma to degree level.

Table 2.3: Characteristics of students in the study by curriculum group

	Characteristic	PBL % (n)		SGL % (n)	
	Married	30	(10)	39	(13)
	Have children	24	(8)	21	(7)
Personal	*Gender male*	*27*	*(9)*	*6*	*(2)*
	1st language not English	21	(7)	33	(11)
		Median	Range	Median	Range
	Age (years)	28	22-55	27	22-50
		PBL % (n)		SGL % (n)	
	Work at Hospital 1	9	(3)	6	(2)
	Work at Hospital 2	18	(6)	24	(8)
	Work At Hospital 3	21	(7)	24	(8)
	Work at Hospital 4	*50*	*(17)*	*39*	*(13)*
Work experience	Work on General Medical or Surgical ward	62	(21)	67	(22)
	Grades F& G	24	(8)	24	(8)
	Staff Nurse Grade E	53	(18)	42	(14)
	Staff nurse grade D	24	(8)	33	(11)
	Ever worked in another hospital	*50*	*(17)*	*30*	*(10)*
	Had a F/T job other than nursing	28	(9)	19	(5)
		Median	Range	Median	Range
	Months at current grade	9	1-84	10	1-120
	Months worked in current ward	18	1-120	18	1-48
	Months worked at current hospital	19	3-192	20	8-204
		PBL % (n)		SGL % (n)	
	Taken an accredited Post registration nursing course	*50*	*(17)*	*77*	*(11)*
Previous Education	*Educated to degree level*	*18*	*(6)*	*6*	*(2)*
	School in the UK	*35*	*(12)*	*21*	*(7)*
	Study leave for all sessions	35	(12)	46	(15)
		Median	Range	Median	Range
	Years since previous formal education	2	0-10	1	0-13

After qualification if a nurse chooses to practice as nurse in the National Health Service she will usually spend a period of time in what are colloquially known as 'General Surgical and Medical Wards'. This is where the bulk of acute in-patient hospital care takes place in the National Health Service. Patients in surgical wards are usually admitted by a Surgeon and their treatment will usually involve a surgical procedure. Patients in medical wards are usually admitted by a Physician and their treatment will not usually involve any surgical procedures. Modern National Health Service management practices lead to patients of both types being admitted to any ward in which there is a bed space. Some of the students in the study worked in a private ward within a public National Health Service hospital that admitted both medical and surgical patients. The distinction between 'General' and 'Specialist' clinical areas is blurred as many Surgeons and Physicians are both 'Generalists' and 'Specialists'. For example an Endocrinologist will admit people with a Myocardial Infarction (heart attack) but will also take referrals from other Physicians for 'specialist' care for Diabetes. For some nurses working in a 'general' clinical area represents a period of consolidation after their training, from which they will move onto to work in a 'specialist' area. For others working in this kind of 'generalist' environment becomes a career choice and for others a place of work that fits most conveniently with their responsibilities outside work.

The UK National Health Service, in London in particular, has traditionally always trained and recruited a large number of nurses from overseas. The Republic of Ireland has been a fruitful recruitment ground for Higher Education Institutions in North East London for many years due to the large Irish community resident in the area. Many of the study participants are from Southern Ireland, which is why the proportion of subjects who were not educated in the UK is higher than the proportion whose first language is not English.

Description of the study participants helps to establish the external validity of the study (in the eyes of the reader). As indicated in table 2.3 the study included students from a wide age range,

a variety of ethnic and cultural backgrounds, different personal circumstances, a variety of educational and professional experience. In this respect the sample appears to be typical of the composition of other nursing programmes in the University and the local NHS full –time nursing workforce. The majority of students were female, single, aged in their early/mid 20s without children. The majority had only worked in nursing, had only ever worked in one hospital, and had done so for less than two years. The majority were educated to diploma level and had recent 'formal' i.e. institutional certificated, educational experience. This profile reflects the fact that the programme included in the study is targeted at junior staff nurses working in 'general' wards in the English National Health Service.

Student motivation and expectations

Student motivation can be characterised as a complex dynamic interplay of orientation, personality, beliefs, cognition, perception, anxiety, and expectations about success, self ability, and learning tasks (McKeachie et al 1986). The literature on the principles of Problem Based Learning pays particular attention to the connection between extrinsic motivation in the form of assessment and 'learning for a job' and intrinsic motivation (the desire to learn) (see for example Schmidt, 1993). Thus the form of motivation assessed for this study was students achievement motivation and in particular the task value component. The student characteristics questionnaire included an open ended question asking students about their reasons for taking the course and their future career plans.

Table 2.4: Self reported student motivation for undertaking course

	PBL %	(n)	SGL %	(n)
To improve knowledge	82	(28)	87	(29)
Came on course to improve practice	50	(17)	36	(12)
Came on course so that can provide more support to colleagues/ students	29	(10)	15	(5)
To improve career prospects	18	(6)	12	(4)
To obtain credits towards degree	6	(2)	33	(11)
Has career plans in nursing	68	(23)	88	(29)

The majority of student's responses consisted of one or two sentences. Analysis of the student's responses produced the categories shown in table 2.4 A high proportion of students responded that they wanted to increase their 'knowledge' though none elaborated on what this meant. Just under half responded that they wanted to 'improve practice'. Another reason given was to obtain credits towards a degree, the difference between the students in the two curricula probably reflecting the difference in the proportion of students in each group who already had a degree. A comparatively small proportion of students made direct reference to improving their employment status as a motivating factor. In response to questions about their future career plans most student responses referred only to the short term goal of gaining a promotion to a higher grade.

Further insight into the students expectations can be constructed from teachers diaries, teaching observations, students comments on evaluations and the feedback obtained from those students who did not complete the course. The main source of this data is the teachers and students in the experimental (PBL) curriculum. The students and teachers in the control (SGL) curriculum made very few comments on these topics. Teachers in the experimental (PBL) curriculum were probably more sensitised to these issues and the students comments were made in the context of Problem Based Learning not meeting their expectations. This does of course mean that this interpretation is more dependent on one group of the students. However, it is suggested that these dispositions were common across these students.

Students appeared to place high value on the 'teacher' as an expert, in possession of 'specialist knowledge'. They appeared to feel that be that they should be 'told' or 'given' this knowledge and that this process of 'telling' constituted teaching. Evidence for this was the frequent complaint made by students in the experimental (PBL) curriculum that they were not being 'taught'. Similarly during discussions of the difficulties they were experiencing and how these could be resolved students in the experimental (PBL) curriculum would state that they needed a

lecture from an 'expert' on a specific 'how to do topic' such as interpreting ECG's. The knowledge that students appeared to value or want was by implication therefore 'external' to them, as such it was seen as 'new' knowledge, the most valuable of which seemed to be procedural or 'how to do' knowledge. Students were motivated by their assessments about which they were anxious to varying degree. This is evident in the way that students expected to have to work between teaching sessions for their assignments but not to prepare for teaching sessions. Students in the experimental (PBL) curriculum frequently remarked that the workload i.e. the work between teaching sessions was too heavy.

Students also appeared to become highly anxious when they felt uncertain about what to do, how to do things, or what standards were required. It appeared as if students desired to be told what to do, how to it, the standard that had to met and then to be tested on whether they did what they were told to do in the way they were supposed to at a high enough standard to 'pass'. The lack of student discussion or comment about the control (SGL) curriculum could be argued to add weight to this interpretation. The students approached the course expecting to be told what to do and how to do it and to be 'told' 'knowledge' or information and as the description below illustrates, in the control (SGL) curriculum this is largely what happened. Student evaluatory comments were therefore almost without exception positive comments about the 'knowledge' of the teachers. By contrast the students in the experimental (PBL) curriculum after initially appearing to be unfazed by the prospect of Problem Based Learning became more and more anxious they perceived they were not being 'taught' 'how to do Problem Based Learning ' or any substantive content.

Harter (1981) proposed five student centred dimensions of the intrinsic motivation to learn in the classroom, challenge, curiosity, mastery, independent judgement, and internal evaluative criteria. Each dimension is a continuum. The challenge dimension refers to student's preference for challenging or easy tasks. The curiosity dimension is anchored at one end by the tendency to

work to satisfy her own interests rather than to please others or obtain good grades. The student's preference for working out problems alone in contrast to relying on the instructor for assistance makes up the mastery dimension. Related to this is the dimension that measures student's belief that she is capable of making judgements about what to do versus being dependent on the instructor for guidance. The last dimension concerns the student's reliance on internal criteria for judging success or failure versus reliance on external criteria e.g. grades. Whilst the students were not formally assessed using Hartman's scales it is argued that the qualitative data presented above suggests that on each one of the dimensions the students in this study err toward the extrinsic end of the motivation continuum. That is their primary sources of motivation lay in extrinsic reward rather than an intrinsic desire to learn. This argument appears to contradict the student's claim that their primary reason for attending was 'to improve knowledge'. It is suggested that whilst this may reflect an aspiration of students, the students' were aware of the normative value of such a claim. Furthermore, it is suggested that the aspiration 'to improve knowledge' does not necessarily mean the same thing as a 'desire to learn'.

Differences in student characteristics between the experimental and control curricula

Randomisation may not result in even distribution of characteristics between the groups therefore it is also important to assess whether the baseline characteristics of the participants in the experimental (PBL) and control (SGL) curricula were similar. In a study design such as that used here where there were no pre-test measures of participants capability, baseline measures become a kind of proxy indicator for the pre intervention ability of both groups. Where baseline characteristics are similar we can be more confident that any differences in outcomes measured after the intervention are a product of the different way in which the Experimental and Control groups were treated during the study. Dissimilarity in baseline characteristics

requires that caution be used when interpreting the results of a study. The difficulty lies in deciding which characteristics are important and what size of difference is likely to affect the outcomes being studies (Altman, 1991). The italicised results in table 2.3 highlight the characteristics on which the students in the Experimental (PBL) and Control (SGL) curricula appeared to differ.

For most of these characteristics, a plausible argument could be made that the difference in baseline characteristic could affect the outcome either in favour of the intervention group or in favour of the control group. For example the higher proportion of people for whom English is not a first language in the control group may result in a lower mean assessment score for this group because these students will have more difficulty in learning and performing written assignments in English. Alternatively, it could be argued that the format of Problem Based Learning with its emphasis on discussion between students as a key means of learning might disadvantage those whose first language is not English. The possible effects of differences in entry characteristics were explored in the data analysis.

Student participation and morale

In both curricula student participation and morale varied between and within groups over the period of the course. This was evident in levels of attendance, participation in classroom activity, quality and amount of work done in preparation for class. Although the teachers commented on student's lack of preparation and participation, the control (SGL) curriculum could and did proceed regardless of levels of student participation in the classroom. Students in the control (SGL) curriculum groups appeared to do little preparatory work or 'studying' between sessions that was not directly related to their assessments. In each of the control (SGL) curriculum groups there were one or two students who would respond to teacher questions

and enter into dialogue with the teacher. There were others who infrequently responded to questions and others who appeared to not respond at all.

> *" The same students still do most of the (student) talking although a few other joined in this session (student name) and (student name) did not say one word to anybody throughout the whole session"*
> (SGL group observation notes 31/01/01)

In three out of the four groups in the control (SGL) curriculum students appeared to be happy and relaxed with each other and with the teacher. In one group however, students appeared to be unhappy and bored on the occasions that they were observed with little interaction between the group members or the group and the teacher.

> *"Students A and B made no notes at all during the session they did not even appear to have a paper and pen. Only student D speaks voluntarily. The body language of students A & B (slouched back in chair looking out of window a lot) suggests lack of interest and boredom"*
> (SGL group observation notes 16/5/2001)

The response of the students to the experimental (PBL) curriculum is discussed more detail below but some preliminary observations about participation and morale are relevant here. The experimental (PBL) curriculum required the students to prepare for classes and required them to learn as a group. This meant that student performance in these areas and their perception of their success or otherwise was a constant focus of the teachers and the students. Notwithstanding this the levels of participation in terms of preparation and classroom interaction did vary between individuals. Interaction between students in the teaching sessions also appeared to be affected by the presence or absence of particular students.

Levels of participation and interaction were linked to student morale, which became a major issue in the groups in the experimental (PBL) curriculum. As described already relations between the students' and their teacher were marked by high levels of anxiety and tension, particularly in the first few months of the programme. These are obvious components of 'morale'. In three out of the four groups in the experimental (PBL) curriculum, teaching meetings, certainly in the first module, tended to be tense affairs with the students appearing

anxious, unhappy and angry with the teacher. In the other group the students still appeared to be anxious but this did not appear to generate as much tension or negativity in the group. Meetings appeared to be more interactive and the students generally more happy and enthusiastic. But even in this group, there were times during most sessions when students appeared to grow bored with the process and to switch off from participating.

> *" The students appear to find the process of setting learning objectives long slow and painful, one student even refers to the process as 'like pulling teeth'. Students appear to drift in and out of the process at times participating and at other times they just look blank"*
>
> (PBL group observation notes 1/3/02)

In the three 'low morale' groups a point seemed to come when the anxiety and tension that had been present became more manageable for the students and in two groups the meetings became more interactive and comfortable. In the remaining group whilst the anxiety and tension levels also fell, this did not seem to result in greater interaction between the group members and the students still appeared to be 'unhappy'.

Programme content and assessment in the study

Programme aims and intended learning outcomes

The programme used in the study is one of a range of Continuing Nursing Education Programmes for which there are a nationally defined curricular frameworks provided by the English National Board for Nursing Midwifery and Health Visiting (now incorporated within the Nursing and Midwifery Council). The framework is broadly defined and institutions apply for accreditation to run a particular programme. In previous years the programme used in this study had been run both separately i.e. as the Advanced Diploma in Medical Nursing and the Advanced Diploma in Surgical Nursing, and together i.e. as the Advanced Diploma in Medical /Surgical Nursing. These changes were largely a result of administrative convenience (i.e. availability of teachers, uptake of the course etc.), but were possible because the cognitive and metacognitive aims and objectives of the programmes are identical even if the substantive content is different. The programmes are provided at level 3 (i.e.3rd year undergraduate) and comprise of three 20-credit modules. Although the programme operates within a modular framework each module is compulsory and the programme is designed as a whole rather than as distinct modules. Students have to successfully complete each module in order to 'pass' the programme.

During the period of this study administratively the two programmes i.e. the Advanced Diploma in Medical Nursing and the Advanced Diploma in Surgical Nursing were run separately. That is the students were not mixed and there were different teachers responsible for running and teaching on the programmes. However, for the purpose of the research study the programmes were treated as one. The teachers involved worked together to design the experimental (PBL) curriculum. The experimental (PBL) and control (SGL) curricula used the

same respective programme handbooks with variation only in labelling, some clinical content and timetabling. The students took the same assessments at the same time.

The programme handbook contains a comparatively lengthy section headed 'Programme Philosophy'. The essentials of this philosophy can be summarised as a belief in education as a process of self-fulfilment and professional development and a belief in the value of experience as a basis for learning through a reflective practice. Some of the issues relevant to the institutional context highlighted above are evident in statements about theory and practice

> *'This programme is an experienced based, modularised course, steeped in practice but underpinned by a wide ranging body of theory ….'*
>
> (A24 Programme handbook 2001)

And about the focus of the course on 'nursing knowledge'

> *'The curriculum holds at its core a strong nursing focus based on the development of nursing autonomy within a multi-disciplinary team. This autonomy requires a new paradigm as to what counts as health care knowledge'.*
>
> (A24 Programme handbook 2001)

According to the handbook the programme approach is informed by Schon's (1987) model of reflective practice realised through using Carper's (1978) framework for nursing knowledge as a model for structured reflection. The programme followed the overall aims and objectives of the University departments Post Registration Nursing Education Provision (see box 2.1). The specific aim is given as *'to foster the development of proficient practitioners in general medical/surgical nursing who will act as role models and resource persons within their clinical area'* (see box 2.2). According to the handbook the learner is encouraged to develop their 'thinking skills', which are described as knowledge, comprehension, application, analysis, synthesis and evaluation. They are also encouraged to examine their 'feelings', which are defined as their underlying values, beliefs, and philosophies, as well as their 'skills'. The programme philosophy aims and objectives applied to both the experimental (PBL) and the control (SGL) curricula. The philosophy aims and objectives as espoused in the programme handbook and by the teachers involved in the

programme are all congruent with the aims of Problem Based Learning and were used as the basis for the design of the experimental (PBL) Curriculum.

Box 2.1: University/ Departmental General aims of Post Registration Education Provision:

- To facilitate the progression of registered practitioners from the level of competence to proficiency in their chosen area of practice or speciality.

Objectives of Post-Registration Education Provision:

Knowledge for Practice:

At the end of a particular programme of study in post-registration nursing, the student will:

- Have moved from a level of clinical competency towards proficiency
- Have developed their skills, knowledge and expertise in the practice area where working.
- Have enhanced their ability to develop and use flexible and innovative approaches to practice.
- Have a comprehensive knowledge of the importance of using evidence based/research findings to inform
- Have developed professional knowledge and confidence to challenge the context and provision of health care

Transferable and Generic Skills:

At the end of a programme of study in post registration nursing, the student will:

- Be able to identify and use resources in order to evaluate new information for practice
- Be able to appraise evidence, critically analyse conflicting theories in order to enhance care.
- Be able to communicate information using logical and rational arguments.
- Have developed skills that enable problem solving and action planning so that practice is developed.
- Have enhanced their ability to work in a team.

Personal Development:

At the end of a programme of study in post registration nursing, the student will:

- With minimal guidance, manage their own learning both academically and professionally, using a wide range of resources
- Have developed an autonomous and reflective approach for continued professional development and life-long learning
- Be able to appreciate the values and beliefs of others

Box 2.2: Advanced diploma in Medical/Surgical Nursing programme aims and learning outcomes

This course will specifically focus on:

- providing the additional education and skills for the medical/surgical nurse to develop their practice to the level of Proficiency.
- enabling practitioners to apply in-depth knowledge of physiology to practice
- Develop skills in setting priorities for care based on in-depth knowledge of the physical, psychological, spiritual and social effects of illness.
- Enable practitioners to develop the use of rehabilitation and health education strategies in their clinical settings.
- enabling practitioners to utilise contemporary approaches to holistic nursing care in their practice.

Module 1: Learning Outcomes:

- Critically examine the role of the medical nurse and its relationship to specialist practice
- Reflect on contemporary approaches to medical/surgical nursing in acute settings
- Apply in-depth knowledge of normal and altered physiology to the care of patients in an acute setting
- Apply models of stress and illness to the practice situation
- Critically appraise relevant concepts and research applying synthesis to nursing practice
- Demonstrate application of theory to clinical practice including the setting of priorities for nursing care

Module 2 Learning Outcomes:

- Examine different systems of care delivery from an organisational, professional and philosophical perspective including primary nursing and integrated care pathways.
- Critically appraise the implications of expanded role issues in clinical practice
- Develop clinical decision making and leadership skills
- Apply theoretical frameworks related to Ethical Dilemmas and the Management of Change to clinical practice
- Develop expertise in the areas of Wound Care and Advanced Life Support
- Synthesise knowledge related to Evidence Based Care and apply it to practice.

*Module 3: Learning Outcomes***:**

- Apply in-depth knowledge of altered physiology to the nursing management of patients needing long term medical/surgical intervention
- Appraise communication skills and increase ability to provide appropriate patient support, counselling and teaching
- Apply the principles of health education to patient teaching
- Master the variety of theoretical perspectives which consider nursing's metaparadigms and conceptual frameworks and evaluate their application to clinical practice
- Critically apply frameworks for rehabilitation for the long term patient, their carers and significant others
- Inspire in others an understanding of the importance of transcultural nursing perspectives in their practice

Assessment and feedback in both curricula

The arrangements for assessment in the programme are described in the programme handbook that is given to students at the beginning of their course. The handbook refers to both formative and summative methods of assessment. Formative methods mentioned include the keeping of reflective practice diaries, the development of an individual learning contract and the development of an individual portfolio. With the exception of the portfolio the students were not required to produce any of these documents in order to 'pass' the course. If any students did prepare these documents they were not shared with the teachers. With regard to the portfolios the teachers simply required the students to bring their portfolio with them on the last day of the course the implication being failure to do so would result in 'not passing'. Therefore the majority of students produced something. In the sense that the portfolio was not shared with the teacher until the last day and the teacher did not give any feedback this would seem to be of little formative use to the student.

Summative methods of assessment used in the programme included an assessment of clinical practice. This consisted of the completion by the students nominated supervisor of a clinical appraisal tool used by the University. The students nominated their own supervisor who had to have attended University training in assessment using the tool. In practice this was usually the students line manager. Teachers required that the student returned the completed forms indicating that their performance was at least competent in order to progress. In practice the students did not seem unduly concerned about this assessment and teachers did not appear to place much importance on the outcome. The summative assessments that appeared to be of most concern to the students and the teachers were the written assignments notionally linked to each module.

The first assignment was a Literature Review with a word limit of 2000 words. For this assignment students were required to conduct a seminar of approximately 30 minutes, produce a handout for other course members, a comprehensive reading list and to write up the topic in the form of a critical literature review. 30% of the marks were given for the presentation, 20% for the handout and 50% for the literature review. The topic matter had to be an issue in clinical practice or area of interest related to 'acute care'. The second assignment was called a Care Study & Supporting Essay and had a word limit of 3000 words. This assignment required students to write a nursing care plan for an individual requiring long term intervention that reflected a particular model of nursing and used evidence/research based rationales for care. In the supporting essay the student was required to explore the tenets of the nursing model chosen and develop a critical evaluation of its general and particular application. The third assignment was called an Action Plan and had a limit of 3000 words. The student was required to identify an area of practice that she would like to improve, produce a critical appraisal of the existing practice and an evidence based action plan detailing how practice could be improved.

Each assignment appeared to have a link to nursing practice. However the intention of the assignments was that students should reflect critically on current practice and how it could be developed or improved using the 'knowledge' /'theory' gained through participating in the programme. This was made clear to students in the instructions for each assignment. The standard university marking schedule used by the teachers reflected this 'level 3' orientation. The marking schedule uses a 20-point scale in which the students have to obtain a mark of 16 or above to pass (1= highest). The students had to pass each assignment in order to progress and in the case of the first assignment both parts of the assignment. The timetable for the submission of the assignments was determined by the university examination timetables. The first assignment had to be submitted in mid December and the second and third assignments very close to each other at the end of April. Teachers encouraged students to submit drafts of assignments for discussion at individual tutorials but this rarely happened unless the teacher insisted on it with a particular student. Students who failed any assignments could resubmit

with a given period. Teachers gave limited comments on the first assignment using standard university forms. This did not appear to happen for the second and third assignments, probably because the students had completed the timetabled teaching sessions before the feedback was available.

Student workload and opportunities for practice

The programme was part-time and the majority of participants (97% n=65) were full time employees. 40% (n=27) were given full study leave by their employers, that is they were given a days study leave for each 'teaching' day. 35% (n=23) were given more than half but not all the 'teaching days' as study leave. 25% (n=17) were given less than half the 'teaching days' as study leave. According to the module handbook each module required 60 hours of 'teaching time' and 150 hours of private study. This is the University stipulated requirement for a level three module. In the control (SGL) curriculum the 'teaching' hours requirement was met on paper at least by scheduling the classes for a full day over for 10 weeks. In practice teaching sessions often did not last a full day and teachers occasionally replaced timetabled sessions with 'self-directed' study time. Based on the teachers records of the time spent teaching in year two the median classroom 'teaching period' was three hours in the experimental (PBL) curriculum and five hours in the control (SGL) curriculum

The programme was designed to fit in a single academic year. Thus students took one module for 10 to 11 weeks with each model following consecutively. What made the course part-time within the University framework was that the students were only doing one module at a time. As the students were working in clinical practice at the same time, it could be argued that they had substantial opportunity to practice what they had learnt. Against this, it could be argued that this time is unsupervised and unsupported in terms of practising or applying newly learnt

knowledge or skills. As will be discussed below an attempt was made in the Problem Based Learning curriculum to build in opportunities for practice.

In order to assess the amount of time that students spent in 'learning related' activity outside of the scheduled classroom teaching sessions students completed a questionnaire that asked them to report what activities they had undertaken in the previous week and how much time they had spent on each. The questionnaire was administered five times over the period of each academic year with a view to obtaining a more reliable estimate. The timing of the administration for each group was selected randomly. Table 2.5 below shows the response rate for each of the five questionnaires by curriculum group.

Table 2.5: Response rates to student workload questionnaires

Workload questionnaire number	Study group (% of entry group) PBL SGL	
1	31 (91)	33 (100)
2	22 (65)	32 (97)
3	22 (65)	31 (94)
4	22 (65)	31 (94)
5	17 (50)	17 (52)

Based on an analysis of all the returned questionnaires the results appear to suggest that students in the experimental (PBL) curriculum visited the library and met their fellow students more often. However, if the analysis is restricted to only those students who completed five questionnaires the median for the two groups is the same. Suggesting that the difference is a student rather than curriculum effect. The amount of study time reported by one student in the control (SGL) curriculum averaged 40 hours per week. This result was a clear outlier being approximately 10 times greater than the nearest comparison in either group. The results from this student were excluded from the comparative analysis of the average amount of time spent on private study per week, the results of which are given in table 2.6 below. The results are

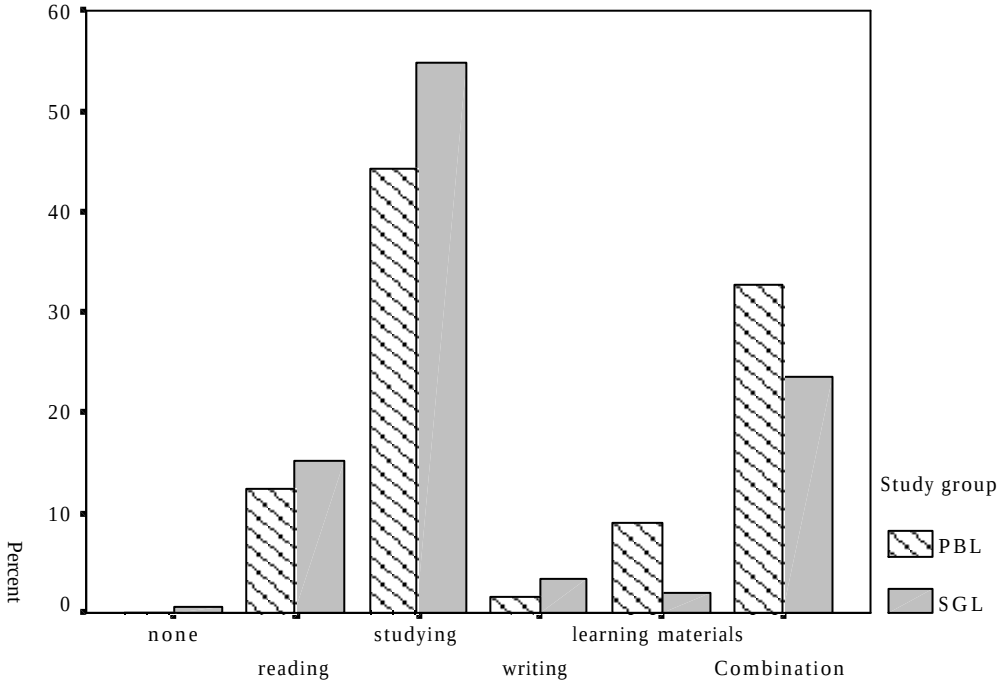
reported for only those students who completed five workload questionnaires. The results suggest that students in the experimental (PBL) curriculum spent more time meeting with colleagues but less time on private study. However, the confidence interval for the difference between the groups does not exclude zero.

Table 2.6 Difference (*d*= effect size) mean amount of time spent in spent in 'private' study per week

	PBL (S.Dev) n=16	SGL (S.Dev) n=16	*d* (95% C.I)
Time meeting with colleagues (minutes)	76 (78)	55 (50)	0.4 (-0.3 to 0.1)
Study time (minutes)	346 (131)	447 (213)	-0.5 (-1.2 to 0.2)

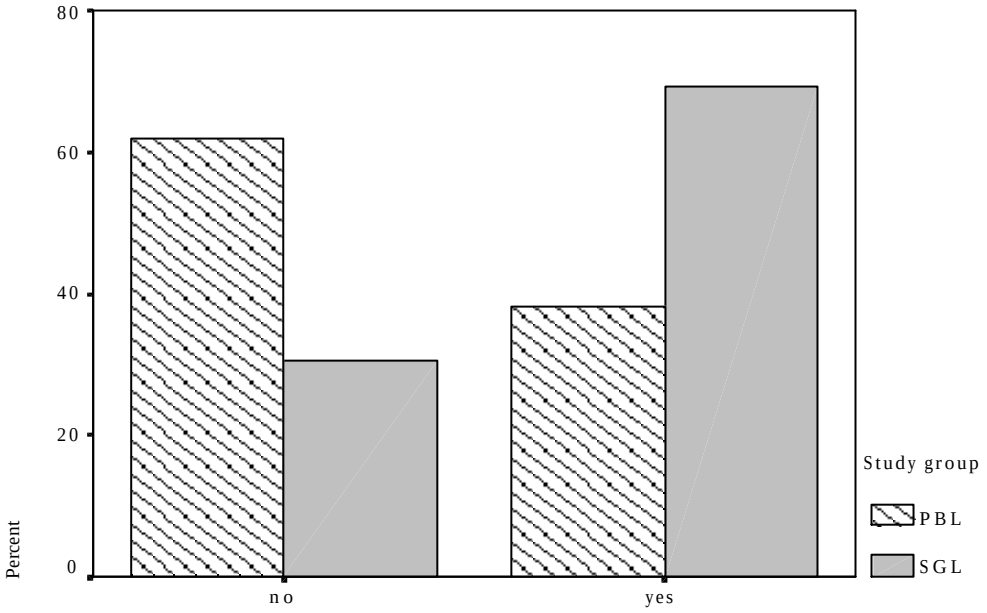
Students were asked about the activities undertaken as private study. Figure 2.5 shows the breakdown of these activities for each curriculum group. The distribution of these activities appears similar in each group. In addition to this student's were asked about the specific purpose of their study activity. The range of responses possible can be collapsed into either for the purposes of assessment or not.

Figure 2.5: Activities undertaken during private study time by curriculum group



As figure 2.6 illustrates students in the experimental (PBL) curriculum were less likely to state that their private study had been for the purposes of assessment (Yes = PBL 43/113 (31%) SGL 99/143 (69%) Pearson's χ^2 24.8 P=0.00). However, the result should be treated with caution as it is based on analysis of all the returned questionnaires and thus the cases are not independent.

Figure 2.6: Purpose of studying for assessment or not



The control (SGL) curriculum in operation

Aims, objectives, philosophies and learning outcomes can be viewed in different ways depending on the view one takes of their purpose (Hussey & Smith, 2002). However, there appeared to be a difference between what Argyris & Schon (1974) call espoused theory (as evidenced in the aims objectives and philosophy) and theory in action (as evidenced in classroom practice). The first evidence of this is also contained in the programme handbook in a section headed 'Outline syllabus content'. A list of 31 topics covered in the programme is given. Only three of the topics listed are concerned with learning, or generic thinking skills/ reflection. Topics relevant to the development of interpersonal or process skills such as teamwork or group work or learning are not mentioned. None of these topics or specific activities to promote them were explicitly timetabled in the control (SGL) curriculum. The timetables should be seen as a record of what did happen rather than as a forward plan. The timetable of the control (SGL) curriculum (see figure 2.7) lists for each session a topic or subject such as *'Total parenteral nutrition and problems with electrolyte imbalance*'. Each week a 'new' topic or subject is listed. The arrangement of topics appears largely pragmatic with little obvious connection or development between sessions. In the observed teaching sessions, the teachers rarely referred to the content of previous or future sessions. These factors combined suggest that in practice the control (SGL) curriculum consists of 'slabs' of subject content related information.

Teaching methods and styles in the control (SGL) curriculum

Despite the claims made in the programme handbook the teaching methods used in the control (SGL) curriculum consisted mainly of 'lecturing' either from the programme teacher or another subject 'specialist' or 'expert'. In the sessions observed the style of lecturing employed by the

teachers in the control (SGL) curriculum varied between teachers and the same teachers' style varied from session to session. In some sessions the style would approximate to that which Behr (1988) describes as 'Exemplary' i.e. using an effective combination of visual and oral modes of presentation. In others, the style appeared to be closer to that which Behr (1988) labels 'Eclectic' i.e. an unplanned mixture.

> *"topic of this session is (session topic) teacher uses a combination of lecturing, OHPs, slides and questions and answers"*
>
> (SGL curriculum teaching observation 31/1/2001)

> *"teacher gave a list of aims and objectives of the session at the beginning but did not seem to have any prepared materials. She left the room for 10 minutes to photocopy an article and then again later to find some OHPs"*
>
> (SGL curriculum teaching observation 13/2/2001)

Approaches to content adopted by the teachers in the control (SGL) curriculum also varied between the teachers and between teaching sessions. On some occasions, the teachers' approach would veer more toward what Pask (1988) describes as a 'Holist' approach, providing a broad overview providing little detail and reaching impulsive conclusions.

> *"the teacher informed the class that the topic for the session is supposed to be (topic). The students were given an exercise on a specific subtopic which the teacher only allowed to continue for a few minutes before starting to lecture on a different sub-topic ……. Later in the session the tutor gives out a handout and appears to be about to introduce another sub topic related to the handout but asks for the time and when told it is 3p.m says "oh better finish there then"..*
>
> (SGL curriculum teaching observation 9/11/2001)

On other occasions the same teachers approach would tend toward what Pask (1988) describes as a 'Serialist' approach, taking a highly detailed narrow focus but providing little in the way of connections between ideas.

> *" The teacher gave the topic for this session as (topic). Most of the session content consisted of the teacher giving a lengthy detailed in depth lecture on biochemistry with only a very short amount of time / activity devoted to consideration of the practical application of this knowledge to the given topic"*
>
> (SGL curriculum observation 7/3/2002)

Figure 2.7: Timetable for a module in the control (SGL) curriculum - Principles and perspectives in Long Term Surgical Nursing NSA 3029

08.01.02	**15.01.02**	**22.01.02**	**29.01.02**	**05.02.02**	**12.02.02**
Introduction to module Learning outcomes Using nursing models/theories in clinical practice Therapeutic nursing	Directed study day – Preparation for group work and assignment on nursing frameworks	am Group feedback on nursing frameworks Pm Total parenteral nutrition and problems with electrolyte imbalance	am Bereavement and care of the dying	am Budgets/Financial management Pm Principles of Health Education	am Practical smoking cessation and risk management strategies pm Managing patients with stomas
19.02.02	**26.02.02**	**19.03.02**	**26.03.02**	**03.04.02** Wednesday	**09.04.02**
Coroner's Court St Pancras	Vascular scanning A+P: carotid, aorta, venous, fistulas (patients present as volunteers) pm Basic pharmacology for surgical practice	Principles of managing change	Practical vascular assessment Dopplers etc pm Bugs! How they affect the surgical patient	am Principles of wound management in the surgical patient Choosing the right dressing Pm Management of chronic pain in the surgical patient	Visit to UKCC misconduct hearings 930 UKCC Portland Place 0207 333 6572

Design and organisation of the experimental (PBL) curriculum

Box 2.3 Generic Problem Based Learning Essentials (Barrows 2003)

- Students must have responsibility for their own learning
- The problem simulations used in problem -based learning must be ill-structured and allow for free enquiry
- Learning should be integrated from a wide range of disciplines or subjects
- Collaboration is essential
- What students learn during their self- directed learning must be applied back to the problem with reanalysis and resolution
- A closing analysis of what has been learned from work with the problem and a discussion of what concepts and principles have been learned is essential
- Self and peer assessment should be carried out at the completion of each problem and at the end of every curricular unit
- The activities carried out in problem based learning must be those valued in the real world
- Student examinations must measure student progress towards the goals of problem based learning
- Problem based learning must be the pedagogical base in the curriculum and not part of a didactic curriculum

Barrows (2003) provides a list of essentials for generic Problem Based Learning (see box 2.3). The Problem Based Learning curriculum in this study was designed with these criteria in mind. The development of the experimental (PBL) curriculum began seven months before the first intake of students onto the programme during the empirical study. Whilst this may seem a relatively short period of time this should be viewed in the context of the fact that this was a one year part time course already in existence. Also, University regulations required that the programme aims, objectives and methods of assessment were required to remain the same for all students taking the programme whichever curriculum was followed. The development of the experimental (PBL) curriculum was undertaken by a Curriculum Sub-Group comprising the Principal Investigator, all the teachers involved in the study, and two senior nurses who had previously completed the programme. The sub group was also able to draw on the expertise of colleagues from other University Nursing departments that had developed Problem Based Learning programmes (McMaster, Southampton and Anglia Polytechnic University).

Woods (1995) nine stage model for the development of Problem Based curricula was used to guide the process of curriculum development (see box 2.4 below). The stages are highly

interrelated. The first stage involved identifying the concepts that were to be explored in the programme. These were identified from the programme handbook and through discussions in the curriculum sub-group. During these discussions it became apparent that there was a substantial hidden curriculum that was not fully reflected in the aims and objectives laid out in the programme handbook.

Box 2.4: Woods' nine-stage model of Problem Based Learning programme development

- Decide how to start
- Visualise the timing and duration of the meetings
- Create the environment for learning the subject knowledge
- Create the environment for the process skills
- Create the environment to develop expertise
- Organise student groups
- Create the resources
- Assess students' performance
- Evaluate programme effectiveness

The concepts identified from the handbook and the hidden curriculum were mapped against the concepts and skills that it is claimed that Problem Based Learning develops. Many of the module objectives were closely linked as were many of the Problem Based Learning concepts/skills (from hereon referred to as skills) and many of the Problem Based Learning concepts were closely related to the module objectives. The exact nature of the relation differs but in most cases the Problem Based Learning skill could be viewed as a component of the module objective. The programme objectives (including aspects of the so called hidden curriculum) and Problem Based Learning skills were linked to develop a curriculum framework organised in terms of five programme meta-aims. The meta-aims are shown as italicised text in box 2.5 below. The programme objectives and Problem Based Learning skills linked under one of the meta-aims are given as an illustration. This programme structure was intended to act as a framework to the tutors and students on their learning journey (Wolff, 2000).

Box 2.5: Experimental (PBL) curriculum framework *(meta aims)*

- *The development of the nurses role in the management of the delivery of care*

Programme Objectives

To critically examine the role of the medical/surgical nurse and its relationship to specialist practice

To explore contemporary approaches to medical/surgical nursing in acute settings

To critically examine different systems of care delivery from an organisational, professional, and philosophical perspective including primary nursing and integrated care pathways

To critically appraise the implications of expanded role issues in clinical practice

PBL skills examples

- Obtaining criteria
- Group skills
- Awareness
- Problem solving
- Reasoning critically & creatively

- *The development of the capability to lead change*
- *The development of the nurses role in promoting health*
- *The Development of clinical expertise*
- *The integration of nursing theory and practice*

The programme ran as three modules and the university operates a two-semester system. This usually means that there is some overlap in modules and/or students may have to attend classes more than once per week. Both of which occurred in the control (SGL) curriculum during the study period. The experimental (PBL) curriculum continued to utilise modules but timetabled meetings so that after the initial preparation days students attended one session per week for a 10-week period with no overlap between modules. Each meeting was scheduled to last 3 hours. In theory, the students the students had the remainder of their ***'study day'*** as free time to work on self-directed learning activities.

The three meeting cycle suggested by Woods (1995) was used as an organising principle for the macro structure of the curriculum. At the first 'Goals meeting' the students read the scenario, explore the issues, prioritise, convert issues to learning objectives, develop criteria, make sure criteria, resources and objectives are consistent, allocate learning tasks and discuss teaching expectations and format. At the 'teaching meeting' each student returns to the group and 'teaches' (gives feedback) on her allocated learning objective. At the 'feedback meeting' the group reviews both individual and group performance and whether their learning outcomes

have been met. Woods (1995) proposes that a fourth 'elaboration' meeting is added in which what has been learnt is applied back to the scenario in order to extend the learning still further.

Figure 2.8 illustrates the organisation of one module from the curriculum. The three-part concept was used with the 'elaboration' activity integrated into the feedback session. The first three 'teaching meetings' were held on three consecutive full days. In these meetings, the principle of the three-part cycle was used within a compressed timescale. After these initial 'preparation days' the three-part cycle was fitted over three teaching meetings spread over 3 weeks. At the first week's goal setting meeting students were presented with a new scenario. Parts two and three of the cycle were integrated into the second and third meeting. The activity associated with Woods' fourth elaboration meeting was carried out at the third meeting where students were encouraged to apply what they had learnt back to the scenario.

Preparation of students for Problem Based Learning

It was recognised that students in the experimental (PBL) curriculum would require additional preparation for Problem Based Learning. The first three days of the experimental (PBL) curriculum were planned as consecutive full day 'teaching sessions' for this purpose. In considering how best to undertake this preparation it was felt important that students should be encouraged to view the development of these capabilities as something that they would work on during course of the whole programme and indeed beyond. It was also recognised that it would be incongruent to attempt to 'teach' these capabilities in a traditional manner and that the 'learning' of Problem Based Learning should therefore be designed as a Problem Based Learning process. In the first year of the experimental curriculum the Problem Based Learning process was followed from day one. Scenarios were used that would generate learning issues relevant to Problem Based Learning, such how do we learn? how to set goals?, how to work as

a group?, how to give feedback etc. As part of the resource list attached to these scenarios students were referred to learning materials that dealt with these issues explicitly. Students were also shown a video 'Problem Based learning in the Tutorial Group', produced by the University of Maastricht. The video was placed in the University library so that student's could view it as often as they wished. The Problem Based Learning tutors were not given explicit directions about how or when these scenarios should be integrated into the first three 'preparation days' but were invited to use their discretion.

At an evaluation of the operation of the curriculum after the first year the teachers felt that this approach had not been as effective as might have been hoped for and that this had some bearing on the levels of students anxiety and discomfort seen. Therefore, the scenarios and materials were formulated into a prescriptive detailed 'teaching plan' that teachers followed during the preparation days. Scenarios were still used but the materials were provided to students in the form of handouts and specific learning activities were given as part of the process. This appeared to work satisfactorily for the teachers and students at the time and reduced student's immediate anxiety. But as discussed elsewhere in the report it did not appear to allay student anxieties and concerns that emerged as the course progressed.

Creating the environment for the development of skills and expertise

As has already been noted the development of process or meta cognitive skills was an explicit programme objective listed in the programme handbook as is the notion of the development of 'expert practice' of which these skills are seen as integral part. The Problem Based Learning 'preparation days' also emphasised the importance of the development of these metacognitive and interpersonal skills as a component of professional expertise. The principle mechanisms for

creating the environment for the development of these skills were the tutorial process and the use of scenarios.

Figure 2.8: Module plan from the experimental (PBL) Curriculum - Medical Nursing programme (2001-2002)

MODULE NSA3025		D 1 28/9	D2 4/10	D3 5/10 am	D4 11/10	D5 18/10	D6 25/10 double session	D8 8/11	D9 15/11	D10 22/11	D11 29/11	D12 6/12
	Structure	a.m. housekeeping p.m. goals meeting s	Teach & feedback	10-45-1 library/ computer visit (searching)	Goals meeting	T&F	a.m. T&F P.m. Goals	T&F	T&F	ECG workshop	Assessed Seminar presentations And module review	Am presentations
	Content	Learning through PBL Identifying and defining problems (Brainstorming) Self assessment Self-directed life-time learning Learning contracts Reflection Identifying learning needs Show PBL video	a.m. feedback – learning contracts & reflection on scenario & PBL	Learning as group – Planning Group skills Chair skills Awareness Interpersonal skills Coping conflict	Physiology Aetiology Prognosis Therapy Psychology Health promotion/ education Sick role assessment skills Pt autonomy		Physiology Aetiology Prognosis Therapy *thrombolysis Pain Mgt Psychology Gender Sick role *Health promotion strategies *Assessment /Observation skills Arrthymia mgt & interpretation Expanded roles Team work				What been learnt Relation to learning contract Things to do different / better Gaps – plans to address	
	Scenario	Joyce is med/surg nurse taking new course, asked to develop learning contract.		Conflict in a clinical team	Sylvia Carter - acute asthma attack		John Cole - Acute angina					

There are different models of the Problem Based Learning tutorial process. The main differences between the models appears to be that in some the process is broken down into more discrete stages and that different language is used. The process referred to in many medical school Problem Based Learning programmes is largely derived from the seven step model developed at Maastricht (see box 2.6 below). This version of the process makes explicit use of hypothetico-deductive terminology and makes no specific mention of the requirement of students to reflect on their development of learning and process skills, although this is a component of the Maastricht model described elsewhere (Dolmans & Schmidt, 2000). This model is featured in the Problem Based Learning video that the students watched.

Box 2.6: The Seven step PBL process used at Maastricht University (Schmidt, 1983)

1) clarifying and agreeing on working definitions of unclear terms/concepts;
2) defining the problem(s), agreeing which phenomena require explanation;
3) analysing components, implications, suggested explanations (through brainstorming) and developing working hypothesis
4) discussing, evaluating and arranging the possible explanations and working hypotheses
5) generating and prioritising learning objectives
6) going away and researching these objectives between tutorials
7) reporting back to the next tutorial, synthesising a comprehensive explanation of the phenomena and reapplying synthesised newly acquired information to the problem(s)

However nursing education in the UK at least tends to place less emphasis on and be less explicit about the use of the hypothetico-deductive approach and greater prominence is given to the role of reflection. This is evident in the decision of some UK nursing departments to adopt of model called Inquiry or Enquiry Based Learning, which it is argued places less emphasis on the Hypothetico –deductive approach and more on understanding and reflection (see Cleverley, 2003 for example). Students were also introduced to a second similar model of the tutorial process. This was modified from an eight step model for Problem Based Learning in nursing education developed by Wolff (2000) (see box 2.7 below).

Box 2.7: The eight tasks of PBL (based on Wolff, 2000)
1.Explore the problem - clarify terms and concepts that are not understandable, create hypotheses, identify issues
2. Identify what you know already that is pertinent
3. Identify what you do not know
4. As a group prioritise the learning needs, set learning goals, and objectives, allocate resources, members identify which task they will do
5. Engage in a self-directed search for knowledge
6. Return to the group, share your new knowledge effectively so that all the group learn the information
7. Apply the knowledge; try to integrate the knowledge acquired into a comprehensive explanation
8. Reflect on what has been learnt and the process of learning

These models of the Problem Based Learning process act as guides for the tutor and students to help them through the learning process. In the first meeting of a cycle, with a new scenario, the students work through steps one to four. Between meetings the students engage in self directed learning. The second and third meetings in a cycle are devoted to getting feedback on what the students have learnt from the research that they have undertaken between the meetings, synthesising and applying this information to the scenario. At the end of each cycle, the group reviews its performance as a learning group and learning goals are identified for improvement. The stages may be worked through sequentially but often the students will move backward and forward between the stages during each cycle as they spend more time thinking and discussing the issues.

Development of scenarios

The curriculum meta-aims and the specific programme objectives were used as a basis for the construction of scenarios. The scenarios were designed collaboratively by the curriculum subgroup and sent for external peer review to colleagues from other Universities already using Problem Based Learning programmes in nursing education. When designing the scenarios the

group used a set of questions derived from the literature on Problem Based Learning and on the design of learning tasks in general (see Box 2.8).

Box 2.8: questions used to facilitate scenario design

- Are the scenarios 'realistic'? Narrative or story line should be usual, common, likely, frequent and should not contain internal contradictions
- Will the scenario generate discussion/ learning issues that is relevant to the curriculum meta-aims and objectives
- Are the learning issues likely to be generated interchangeable across the different working environments of students (i.e. different wards, hospitals)
- Is there sufficient descriptive and contextual information in the scenario to allow exploration by the students
- Is the scenario likely to be relevant to the students
- Is the scenario likely to prove interesting to the students
- Is the scenario challenging for the student at their current level of knowledge
- Does the scenario integrate different subjects/ disciplines
- Does the scenario promote self –directed learning – can the student ask and follow-up their own questions
- Is the scenario 'doable' i.e. can the learning task be completed in the time available
- Does the scenario contain sufficient information to guide the student's identification of appropriate learning resources?

In the literature on Problem Based Learning the terms 'Problem', 'Trigger' or 'Scenario' are used to refer to the material presented to students for initiating a specific learning cycle. Often these terms are used interchangeably even when in practice there appear to be significant differences in the material presented. The use of the term 'Problem' as in 'Problem' Based Learning refers to a problem in the cognitive sense. Based on studies of expert decision making Elstein and colleagues (1978) set out the distinction between well-defined, moderately defined and ill-defined problems. For a problem to be well defined there must be one clearly preferable solution and a small change in the problem would result in only a small change in the solution. Where more than one potentially acceptable solution exists the problem is described as 'moderately well defined'. For ill-defined problems, there may be no solution or there may be one solution and small changes in the problem will require large changes in the solution. The 'authentic' Problem Based Learning approach described by Barrow's uses ill-defined problems to simulate the conditions that occur in the real environment. 'Problems' by this definition are therefore

situations that challenge existing knowledge and expertise and invoke the hypothetico-deductive process (Myers Kelson & Distlehorst, 2000).

Linked to this is the research on medical and clinical reasoning that has shown the importance of organisation and memory structure to explain differences among novices and experts (see for example Benner, 1984; Elstein et al, 1978). One way of expressing this organisation is the notion of memory structures called semantic networks. A semantic network is an elaborate set of meaningful connections among abstract concepts and or specific experiences. The acquisition of expertise in an area can be characterised by the development of rich semantic networks that are adapted to the tasks in that domain of expertise. Initially a novice has related concepts but only a few to work with. With experience and education new concepts and concrete examples are added to the network and new stronger richer connections are made between existing concepts and examples (Regehr & Norman, 1996).

It was felt important to try to avoid giving students the impression that the 'task' of Problem Based Learning in this curriculum was solely to solve 'a' or 'the' 'problem' in any presented material. The term 'Scenario' was therefore used to refer to the material presented to students. The Scenario's were intended to play at least three roles. Firstly discussion of the scenarios serves to encourage students to activate relevant prior knowledge (Schmidt & Moust, 2000) and for this reason scenarios were designed to be 'familiar' to students (Soppe et al, 2003). Secondly, to stimulate students interest and thus their intrinsic motivation to learn. Thirdly, to set a context for the learning of knowledge which is similar to that in which future use of the knowledge will be required (Schmidt & Moust, 2000). For example from the scenario given in box 2.9 below students in the year one 'medical nursing ' experimental (PBL) curriculum identified four broad learning issues. 'What are effective styles of leadership?' 'How best to implement change?' 'What are current NHS & UKCC guidelines on patient documentation?' 'What is clinical supervision and preceptorship?'

Box 2.9: example of scenario from the Problem Based Learning curriculum: Nurse Sue Downs

Joy Chen is the newly appointed F grade in charge of Blue team. She has been in post 2 months. Sue Downs is a D Grade nurse in Joy's team and has worked on the ward part-time for 10 years. Sue is popular with other members of staff and with the patients. Joy feels that the team's documentation of patient assessments, care plans and evaluation could be improved. She perceives that when Sue is the named nurse documentation is particularly poor. She also notices that when Sue has been looking after a group of patients on a shift she frequently leaves work incomplete. Sue is quite open about this often reporting that 'she has not had time to do such and such' in the handover meeting. Although nobody complains about this Joy feels it may cause resentment amongst the other team members who have to 'do her work for her'.

Resources:

Adair J (1986) Effective team building. London. Pan books
Douglass, LM (1992) The effective nurse - leader and manager 4th ed. St. Louis : Mosby-Year Book .
Sullivan, M P (1990) Nursing leadership and management . Springhouse (Pa.) : Springhouse Corp.
Tappen, R M (1995) Nursing leadership and management - concepts and practice. 3rd ed. Philadelphia. Davis
Driscoll J (2000) Practising Clinical Supervision: A reflective approach. Edinburgh: Bailliere Tindall.

Core Concepts: Nurse Sue Downs

- Nursing Role: Aesthetics: *Management of change, Leading & motivating a Team, Role modelling
- Ethics: Leading a team, Dealing with staff
- Professional Role: *The nurse as a change agent, *Reflective practice, risk management, Clinical supervision, Leadership, performance management
- Sociology: Role Theory, Management of Change theories, Organisational and occupational culture,
- Education/ Psychology: Motivation theory, theories of reasoned action, learning styles, learning theory

A scenario consisted of a short narrative, additional information pertinent to 'the case' and a directory of further resources. The list of 'core concepts' derived from the programme aims and objectives were only given to the tutors to act as a guide to the relationship between a particular scenario and the overall aims and objectives of the programme. Ten scenarios were used in the experimental (PBL) curriculum (see box 2.10 below for a list of all the scenarios used). The majority of scenarios were identical for the surgical and medical groups with the exception of the scenarios that had a specifically 'clinical' orientation. A review of the scenarios after the first year

of the curriculum suggested that only minor modification of technical detail in two scenarios was required and the order in which the scenarios were presented was altered slightly.

Box 2.10: List of scenarios in the experimental (PBL) curriculum

Generic scenarios

- Nurse starting a new course required to develop leaning goals through a learning contract
- Conflict between members of a clinical team
- New ward manager wants to introduce a philosophy and model of nursing care in a non-consensus environment
- A member of the nursing team is under-performing
- Manager of a ward is told to 'do something' about Evidence Based Practice

Clinical Scenarios – Medical Nursing

- Young women with repeated admissions for acute asthma attacks
- Middle aged self –employed man admitted with acute Angina
- Same man has a cardiac arrest
- Young Sikh mother terminally ill with Breast cancer metastases
- Elderly man admitted with stroke and dense left Hemiplegia

Clinical Scenarios Surgical Nursing

- Young women admitted for formation of temporary colostomy due to Crohn's disease
- Middle aged self- employed man admitted with acute cholecystitis
- Same man has a cardiac arrest post-surgery
- Elderly lady who had surgery after a fractured neck of femur
- Afro- Caribbean man who has a below knee amputation due to poorly controlled Diabetes

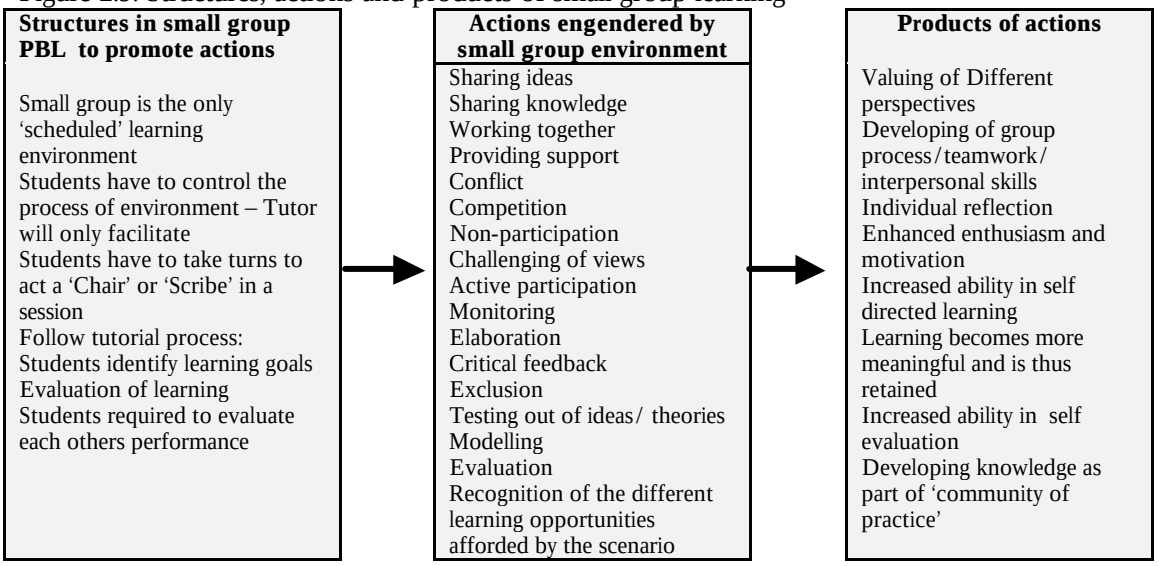
Organisation of the small groups in the Experimental (PBL) curriculum

The small group is an integral part of the Problem Based Learning approach, used consciously and conscientiously to achieve the learning outcomes (Benson et al, 2001). It is argued that purposefully designed and successful small group learning facilitates learning through the development of a learning environment that supports and promotes both cognitive and metacognitive development. Theoretical support for small group learning can be found in a variety of educational literature with different emphasis such as Adult Learning (Brookfield & Preskill, 1995), Group Processing (Sampson & Marthas, 1990), Collaborative Learning (Bruffee, 1993), Learning Organizations (Watkins & Marsick, 1993), Collaborative Knowledge Building (Bereiter & Scardamalia M, 2000), Cooperative Learning (Cohen, 1994), Constructivism (Tynjala,

1999), Situated Learning in Communities of Practice (Wenger E, 1998) and Cognitive Processing (Schmidt & Moust, 2000) to list but a few.

The links between the structures of small group learning emphasised in Problem Based Learning and the actions and learning outcomes that it is claimed result from these actions are illustrated in figure 2.9. Implicit in the design of the Problem Based Learning small group is the idea that many of these 'positive actions' e.g. co-operation do not just happen by themselves, whereas many of the 'negative actions' e.g. conflict are a routine and inevitable part of working in a group. The 'structures' in small group Problem Based Learning, along with the tutorial process and the use of scenarios, help the students to learn how to learn in groups and learn how to anticipate, prevent, cope and deal with the difficulties that they will experience working in this way. This is not to say that these structures are present in the organisation of all Problem Based Learning small group learning environments.

Figure 2.9: Structures, actions and products of small group learning



The small group meeting was the only 'scheduled' learning experience for the students in the experimental (PBL) curriculum. The small group process used included the requirement that at each session a different student was required to facilitate or chair the session. Firstly this reinforced the message that students were required to take responsibility for the learning process and for their function as a group. Secondly, it is argued that 'facilitation skills' are an important

part of the professional repertoire. Taking on the role of facilitator in a supportive environment helps students to practice and develop these skills (Benson et al, 2001). Students were given written information about how to chair and students were asked to give feedback to the chair on her performance at the end of each session. Each week a different student was asked to 'scribe' that is to make notes of the discussion during the session on a flip chart. The tutors 'modelled' the 'role of 'chair' and 'scribe' process on several occasions and these too were subject to feedback at the end of each session. The students were provided with guidelines for giving feedback and introduced to various tools for this purpose. At the end of each session students were asked to evaluate their performance as a learning group during the session and to identify goals for improvement.

There are differences of opinion about the ideal size for a Problem Based Learning group, but it is argued that the development of skills for communication, the development of knowledge and collaboration are best fostered in a small learning group (Benson et al, 2001). One of the reasons for selecting the programme used in this study was that based on past enrolments it was anticipated that it would be possible to create small groups. The number of students enrolled for the programme and the randomisation process determined the size of the groups. The randomisation process was carried out in blocks of ten to maintain equal group sizes. The actual number of starting students and completing students in each group is given in table 2.7 below. The surgical programme under –recruited in both years of the study resulting in smaller groups than anticipated. The medical groups in the experimental (PBL) curriculum started larger than the five to seven recommended by Myers Kelson & Distlehorst (2000) but due to drop out fell within that range by the end of the course. The surgical groups in the experimental (PBL) curriculum started within the recommended range but in the first year at least became, it could be argued, too small.

Table 2.7: Group size at the start and (finish) of the programme

	Medical PBL	Surgical PBL	Medical SGL	Surgical SGL
Year 1	9 (4)	7 (3)	10 (9)	5 (4)
Year 2	11(7)	7 (5)	12 (12)	6 (6)

Role of the teacher in the Experimental (PBL) curriculum

The role of the tutor is argued to be of crucial importance in Problem Based Learning (Wolff & Rideout E, 2001). The general conception of the role of the teacher in Problem Based Learning is embedded in the conceptions of knowledge and its development that underpin Problem Based Learning. For some the philosophy of Problem Based Learning tutoring can be traced back to Socrates (Johnson & Finucane, 2000). The Socratic notion of 'the teacher as midwife to the ideas of students' runs through most conceptions of the role that, in one form or another, the take the view that the teacher is a more knowledgeable member of the same social community as the student. (Rideout & Carpio, 2001).

The issue of just how knowledgeable or knowledgeable about what, seems to have exercised the Problem Based Learning community greatly judging by the volume of literature that discusses or reports on the advantages and disadvantages of 'expert' tutors compared with 'non-expert' tutors. This issue is discussed in the section on 'teachers experience' above and will not be repeated here. However, there are two points emphasised in the Problem Based Learning literature that are linked to this issue that should be noted. Firstly that following from Dewey there is no support for so called 'context free facilitation' in the Problem Based Learning literature (Margetson, 1993). Secondly, whether the tutor is a subject expert or not there is universal agreement that Problem Based Learning teachers should not provide the students with what *the teacher* feels is the information the students need i.e. *'the answers'* (Barrows, 2000).

The teaching role of the 'more knowledgeable member of the community' is conceptualised in both socio-cultural and cognitive terms. Socio-cultural approaches emphasise the teacher's role in enculturating the learner into the specific community of practice through for example internalisation of the language, attitudes and values of the community (Bailey et al, 2003). Cognitive approaches emphasise the teacher's role as facilitator of cognitive development in the knowledge and skills of the profession (Schmidt & Moust, 2000). Whichever conception is

emphasised the primary use of the teachers subject knowledge is the designing and provision of the problem experiences (or scenarios as they have been termed here) (Barrows, 2000). From a different perspective, Savin-Baden (2000) describes the role of educational institutions as creating 'disjunction' for the students (recognition of gaps or challenges to existing knowledge) and 'opportunities for transformation' (learning).

Box 2.11: Teaching techniques used by the PBL teacher (after Wolff, 2000)

Communicative actions	Role personae
• Staying silent • Probing questions: e.g. Why, What do you mean, What does that mean? • Reflecting questions: How does this idea help you • Involvement questions e.g. who else has ideas on this? • Physical positioning in group • Educational diagnosis questions e.g. how do you feel about the way you formulated your ideas ? • Stimulating interest • Decreasing challenge where there are signs of boredom or 'over challenge' • Helping students to address issues with interpersonal dynamics e.g. by asking questions about dysfunctional group behaviours	• Learner • Creator • Director • Challenger • Evaluator • Negotiator • Modeller • Designer • Facilitator • Supporter

In the classroom the Problem Based Learning teacher employs their knowledge of the 'subject' area to support the processes of cognitive or meta-cognitive development and/or enculturation. The Problem Based Learning literature suggests a number of techniques that the Problem Based Learning teacher may adopt in their interaction with students. These techniques include the adoption of particular role personae and forms of communicative action (see box 2.11 above). There appears to be some discrepancy in the conception of the role as outlined above and the and the idea put forward by Barrow's (2000) that teachers should act as 'consultants' or resources to be used in the same way as text books for example. Schmidt and Moust (2000) also argue that the teacher should provide answers where the students are in genuine difficulty. This may be linked to what Margetson (1998) describes as the 'Convenient Peg' conception of Problem Based Learning where the basic learning task is still viewed as 'knowledge acquisition', i.e. models of Problem Based Learning that Savin-Baden (Savin-Baden, 2000) labels for 'Epistemological

Competence' or 'Professional Action'. The techniques and role personae listed in box 2.11 were those recommended to the Problem Based Learning tutors in this study as part of their preparation for the role. The role of the Problem Based Learning tutor in this study appears to be congruent with that described in Savin-Baden's (2000) model of Problem Based Learning for 'Transdisciplinary Learning' or 'Critical Contestability' i.e. as an orchestrator of learning opportunities, a commentator, challenger and decoder.

Preparation of the teachers for the role of Problem Based Learning tutor

The Problem Based Learning tutors were faculty members who volunteered for the role. A frequent complaint of the tutors during the course of the study was that they had insufficient time and preparation for their role as Problem Based Learning tutors. The preparation of the curriculum and tutors began formally some five months before they 'started' tutoring. The departmental management team was supportive of these teachers request for relief from other responsibilities in order to take part in the study but it is unclear how much practical use this was to the teachers. However against this the tutors were senior, experienced, qualified nurse educators who when they became involved in the study stated that they were already using Problem Based Learning in their teaching.

The task of tutor preparation was therefore a delicate process of resocialisation into a role which if not entirely new to the teachers involved, ultimately proved a challenge to their values and beliefs about teaching as well as their self image as 'expert nurse teachers'. All four of the resocialisation processes outlined by Fallon and Pomfret (1977) were used over the period of the design and implementation of the experimental (PBL) curriculum. The principal emphasis was placed on involvement of the tutors in the design of the Experimental (PBL) curriculum as this facilitated their exposure to the key ideas and literature in the Problem Based Learning field. The tutors participated in a number of external workshops on Problem Based Learning run by 'experts' from other Higher Education Institutions and observed Problem Based Learning

sessions at another University. When the first experimental (PBL) curriculum commenced the Principal Investigator sat in on a number of 'teaching' sessions and offered advice and feedback to the tutors but this declined as the tutors became more confident. The teachers and the Principal Investigator continued to meet as a group throughout the period of the study to offer advice and support to each other. During the first year of the study these sessions also included a senior member of the faculty teaching staff who was completely independent of the research and teaching programmes and whose 'supervisory' role was to help the teachers with the some of the challenges they experienced during the project.

Teachers performance in the experimental (PBL) curriculum

Reference has been made elsewhere to teacher emotional and cultural responses to the experimental (PBL) curriculum. The focus here is on their 'performance' in terms of the role conceptions and techniques described above. The amount of concentration, skill, and creativity required to sustain interactivity using the techniques outlined above should not be underestimated, particularly in the face of student anxiety, hostility and constant attempts by the students to get the teachers to revert to a more traditional lecture format. The Problem Based Learning teachers appeared to find the role more difficult than they had expected and not surprisingly their level of performance in this regard varied from week to week.

As indicated by the interaction analysis (see pages 141-143) teachers in the experimental (PBL) curriculum spent a considerably smaller amount of time 'lecturing' students i.e. they spent a comparatively small amount of time giving them information that the teacher felt was needed. However, this still happened during sessions and seemed to be linked to the teacher's need to establish or reinforce their status as 'a more knowledgeable member of the community'.

> *"Tutor 'lecturing' for about 20 minutes 'drilling' about ECG's and running through different rhythms using ECG monitor'*
>
> (Teaching Observation PBL session 14/11/00)

Although the Problem Based Learning teachers employed the full range of techniques/ role personae outlined in box 2.10 they appeared to employ few variations in each technique and seemed to miss opportunities to probe student's knowledge.

> *The students 'covered' the topic of change management in about 50 minutes most of which had been spent with them recounting 'stories' from their own workplace. The students concluded that change theory was not all that relevant, one student stating '"its all about the quality of the individual change agent'. The tutor did not question or challenge this*
>
> (Observation notes PBL session 29/03/01)

Both Problem Based Learning teachers reported frustration and difficulty in 'getting the students to go deeper' i.e. to deepen their understanding. To an extent this needs to be viewed in the light of the students expectations and motivation discussed above.

Students response in the experimental (PBL) curriculum

Generally the students did not appear to meet the tutor's expectations with regard to the development of their ability or willingness to take responsibility for the learning process. The observations of the Problem Based Learning groups during the study also identified a number of 'actions' that did not appear to be fully or independently (i.e. by the students) activated by the combined structures of the scenario, tutorial and group process.

Tutors indicated that attendance rates were not what they expected. The Problem Based Learning tutors diaries listed 12 comments about poor attendance. However this impression does not appear to be born out by the attendance data which showed the average attendance in the experimental (PBL) curriculum to be 88% (28 out of 32 sessions) and in the control (SGL) curriculum 87%. The lowest rates of attendance for those who completed the programme in each course were 84% (27 out of 32 sessions) and 73 % respectively.

In general students struggled with the non-directivity of the scenarios. Their preference was to move into a problem-solving pattern immediately – identifying *'a'* patient problem or question from the scenario that needed to be 'solved'. In general this problem solving pattern was characterised by a tendency toward what is referred to in the literature on Problem Solving as 'premature closure' i.e. settling on the first question and answer that came to mind (Kahney, 1993). Perhaps reflecting the general tendency toward 'coverage' both on the part of the teachers and the students noted earlier, once questions or learning objectives had been identified students then divided them up between each other for action. Students would go away and refer to a small number of textual information resources (usually textbooks). There were exceptions such as the student who visited the local Department of Social Security Office to find out about the benefit entitlements of self employed people or another student who searched the internet and compiled a summary of evidence based wound care sites. The feedback sessions usually took the form of a student reading out from a text in a mini– lecture format.

Thus group interaction in respect of the tutorial process was largely limited to the 'mini-lecture' type identified by Hadwin (1996). Whilst this is recognised as a kind of first 'stage' in the student's development in the Problem Based Learning literature (Benson et al, 2001), according to Hadwin (1996) it is the least effective at promoting self regulation and fostering critical thinking. There was progress within this type. In sessions at the beginning of the programme this 'mini-lecture' often consisted of reading from a photocopied article and students gradually progressed to preparing and reading from their own summaries. But even where this happened the focus of the student's feedback appeared to be providing an answer rather than exploration of an issue. Some students did progress to using visual aids and other feedback formats such as role-play and video but these were exceptions.

The process of developing a more interactive and complex feedback process within the group is in part facilitated by the engagement of the students in self-reflection, and feedback. The students appeared reluctant to investigate completely new topics i.e. things that they perceived that they knew nothing about. The nature of these subjects would vary from group to group. The typical

strategy for avoiding engagement with these topic areas (which were often raised by the tutor) was to state that this was outside the remit of their role as nurses. This was accompanied by what appeared to be a reluctance to reflect upon and critique their own and others knowledge, at least not in the public space of the classroom. As the interaction analysis below indicates there was a good deal of inter–students dialogue in the experimental (PBL) curriculum, however much of the discussion would be of the type 'this is what we do in our hospital or in my ward'. There would be little attempt to explore the rationale behind these statements and even where differences were apparent few attempts to reflect upon the reasons why this might be the case.

A underlying trend for improvement in the students engagement with the scenarios, tutorials and group processes as the programme progressed is detectable in the Problem Based Learning tutors observations.

"Students have only superficially looked at the literature…… students could not discuss or analyse"
(PBL tutor X year 1 module 1 diary)

" good exploration of issues in discussion ….. started to self-evaluate and critique"
(PBL tutor X year 1, module 3 diary)

However even within this generally progressive trend the Problem Based Learning tutors noted variation in student performance from week to week right up to the last weeks of the programme

"they have regressed, there is still a lack of preparation even in module three"
(PBL tutor X year 2, module 3, diary)

Based on the classroom observations and teachers accounts the groups in the experimental (PBL) curriculum demonstrated recognisable characteristics of progress through the stages of group development. Table 2.8 below shows summarises the position of each of the groups in the experimental (PBL) curriculum with regard to the 'Properties of the Group' and 'Stage of Development' given by Wolff (2000).

Table 2.8 Experimental (PBL) curriculum groups progress in the various properties of group development

Properties of the group	Early Phase: trust	Early Middle stage of self-assertion	Late middle stage of Intimacy	Early mature stage of Identity	Late Mature Stage of Equilibrium
Norms and standards			Works to own standards, Less testing behaviour, Suspicious of deviant behaviour Groups1:1 1:2, 2:2, 1:1	Less pre-occupied with rules & standards, Less attention directed toward deviant behaviour 2:1	
Affect		Some hostility among members Group 1:2	More relaxed, reduced hostility, begin to enjoy Groups 1:1, 1:2	High level of interaction, Real feelings and affection shared Group 2:1	
Tasks and functions			Task roles more apparent – work level higher but still variable All groups		
Leadership		Pairing & subgrouping takes place Group 1:2	Leadership hierarchy apparent Group 2:2	Leadership shifts depending on situation Groups 1:1, 2:1	
Conflict resolution		Issue of Trust not yet resolved conflicts not openly acknowledged All groups			
Cohesion		Members continue to test each other Group 1:2	Some cohesion more trust evident, Beginning of group feeling Groups 1:1, 2:1, 2:2		

The stage of group development reached by the groups was different for each of the properties. The pace of progress in group development also varied between groups. Group 2:1 moved through the stages to their 'highest' level early in the programme but then became stuck and did not progress to the 'highest' Late Mature Stage. The other groups stayed in the Early 'stage of Trust' phase for a considerable period. For all three of these groups, the apparent catalyst for their progression was the departure of one or more of their group members from the programme. This phenomenon and the failure of all groups to progress to the highest stage may have been linked to the fact that students in all the groups were extremely reluctant to engage formally in group process activities such as evaluation and feedback. Students appeared to be

tacitly aware of some of these dynamic issues as manifested in the ways particular individuals were isolated or ignored at various points in time. However the students would or could not discuss these kind of issues despite being given frequent opportunities to do so. Very often group dynamics improved dramatically when particular individuals either quit the programme or were absent from a session. By the end of the programme, three out of the four groups in the experimental (PBL) curriculum appeared to have become a cohesive supportive unit and to have developed a strong sense of group identity.

Exploring differences between the two curricula

Classroom interaction in the experimental and control curricula

It could be argued that the commitment in the control (SGL) curriculum to develop the broader learning and process skills listed in the programme handbook was being met through the learning strategies used rather than the curriculum content. The programme handbook states that a variety of 'Learning Strategies' will be used in the control (SGL) curriculum of which 'lectures from specialists' is only one of ten listed approaches. As part of the process evaluation five teaching observations were undertaken for each group in each year of the study out of a total number of sessions for each group of between 27 and 33 sessions per year depending on the curriculum/group. The sessions were selected randomly from the timetable and the teacher asked/informed only the day before. In each teaching observation three 15-minute sections of classroom activity were audiotaped. Each 15-minute section was taped at roughly the same time, after half an hour, roughly midway and towards the end of a session. The impression gained by the observer was that the predominant mode of classroom interaction in control (SGL) curriculum was of a kind of teacher talk that can be characterised as 'lecturing'.

Flanders (1970) Interaction Analysis is a system of classroom interaction analysis. The system in its original and modified forms have been used extensively in classroom observation studies (Wragg, 1999). It has also been used in the study of differences between expert and non-expert Problem Based Learning tutors at University of Michigan Medical School (Davis et al, 1992). The Flanders Interaction Categories (FIAC) consist of 10 categories of communication which are said to be inclusive of all communication possibilities (see box 2.12 below).

There are seven categories used when the teacher is talking and two when the pupil is talking. Because the system is totally inclusive, coding at a constant rate allows calculation of the proportion of time spent in one or more categories. To illustrate the difference between the

experimental (PBL) and control (SGL) curricula one session from each teacher in each curriculum was analysed. The investigator listened to and coded a six minute segment of interaction from each of the three 15 minute taped sections. The coding process requires that a 'tally' i.e. a category code, is made every three seconds, resulting in a total of 360 tallies from the 18 minutes analysed from each teaching session. These codes were then analysed using SPSS.

Box 2.12: Flanders' Interaction Analysis Categories (FIAC)		
Teacher – talk	Response	**1.Accepts Feeling**. Accepts and clarifies an attitude or the feeling tone of a pupil in a non threatening manner
		2. Praises or encourages. Praises or encourages pupil action or behaviour. Jokes that release tension, but not at the expense of another individual; nodding head, saying um, hmm or 'go on' are included
	Initiation	**3. Accepts or uses ideas of pupils**. Clarifying, building or developing ideas suggested by a pupil. Teachers' extensions of pupil ideas are included but as teacher brings more of his own ideas into play, shift to category five.
		4. Asks questions. Asking a question about content or procedures; based on teacher ideas , with the intent that the pupil will answer
		5. Lecturing. Giving facts or opinions about content or procedures; expressing his own ideas , giving his own explanation or citing an authority other than a pupil
		6. Giving directions. Directions, commands or orders to which a student is expected to comply
		7. Criticising or justifying authority. Statements intended to change pupil behaviour from non acceptable to acceptable pattern ; bawling someone out; stating why the teacher is doing what he is doing ; extreme self-reference
Pupil Talk	Response	**8. Pupil-talk – response**. Talk by pupils in response to teacher. Teacher initiates the contact or solicits pupil statement or structures the situation. Freedom to express own ideas is limited
	Initiation	**9. Pupil-talk – initiation**. Talk by pupils that they initiate. Expressing own ideas; initiating a new topic; freedom to develop opinions and a line of thought, like asking thoughtful questions; going beyond the existing structure.
Silence		**10. Silence or confusion**. Pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.

Figures 2.10 to 2.14 illustrate the patterns of interaction for each teacher in terms of the proportion of the total observed time spent on each form of interaction.. There was some variation in the patterns of interaction in each group from week to week but the distinctive patterns seen in these examples is consistent with the observations field notes for all the sessions observed.

Figure 2.10: Classroom Interaction: experimental (PBL) curriculum - Teacher Y 22/2/2001

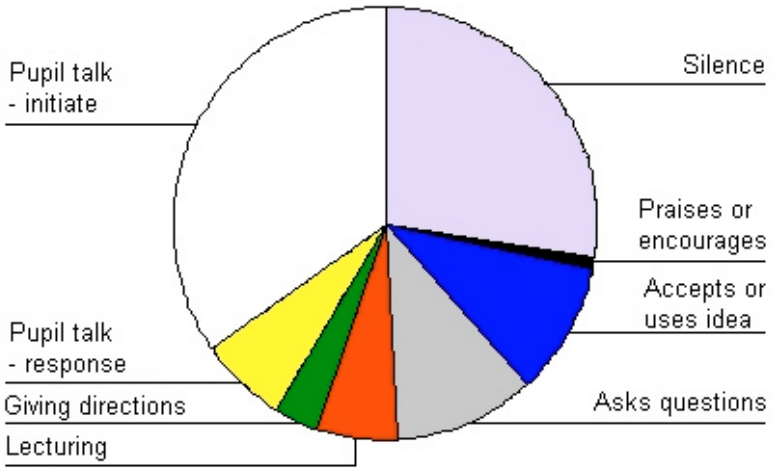
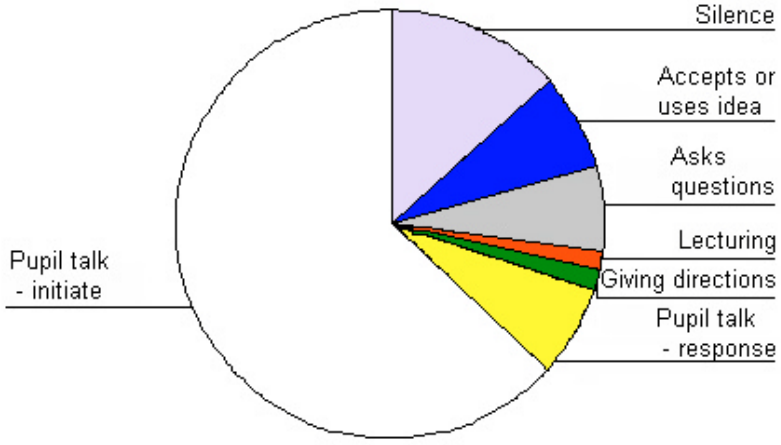


Figure 2.11 Classroom Interaction: Experimental (PBL) Curriculum - Teacher X 21/3/2002



The proportion of pupil initiated talk is greater in one of the Problem Based Learning groups and the proportion of time spent in silence greater in the other. On the days of the observations

were four students present in one group and five in the other suggesting the different interaction patterns was not due to differences in group size. The students in Teacher Y's group did appear generally quieter and less likely to initiate discussion with each other whilst the students in this Teacher X's group appeared to be more ready to voice their views and opinions. The amount of teacher centred interaction in the experimental (PBL) curriculum is slightly higher than that found in studies of 'expert' Problem Based Learning tutors (Hmelo, 2003).

Figure 2.12: Classroom Interaction: Control (SGL) curriculum teacher 1 31/1/2002

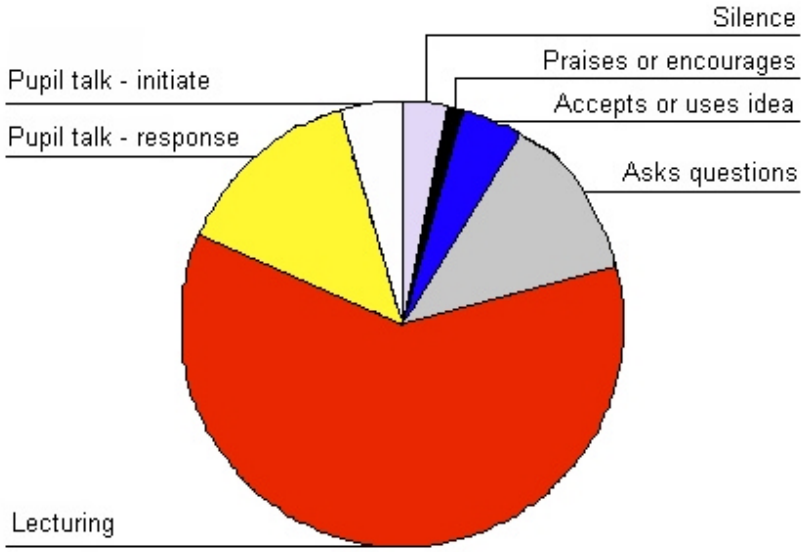
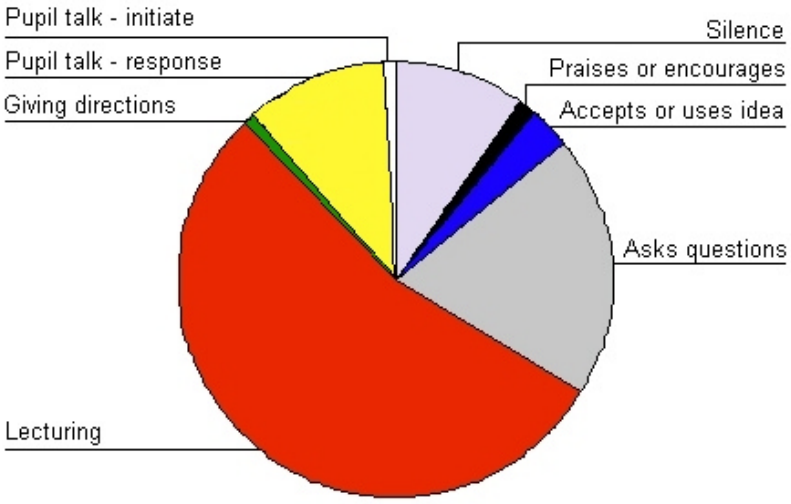


Figure 2.13: Classroom Interaction: Control (SGL) curriculum teacher 2 20/3/2001



In one of the control (SGL) curriculum groups (figure 2.14), pupil initiated talk comprised a much larger proportion of interaction than in the other two groups in the control (SGL) curriculum groups. This difference whilst perhaps less marked was consistent across the sessions observed.

Fig 2.14 Classroom Interaction: Control (SGL) curriculum, teacher 3 16/4/2002

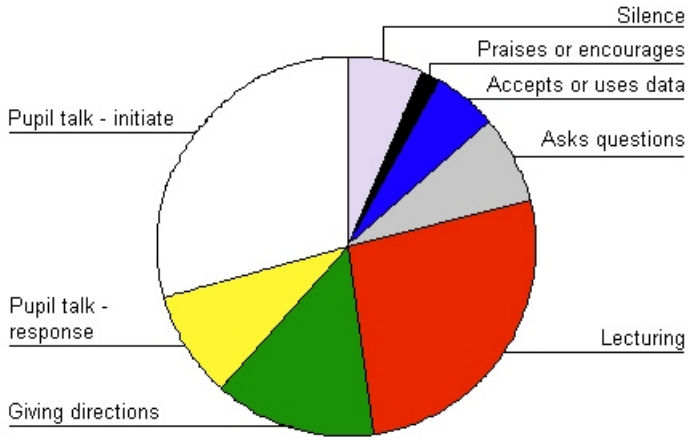
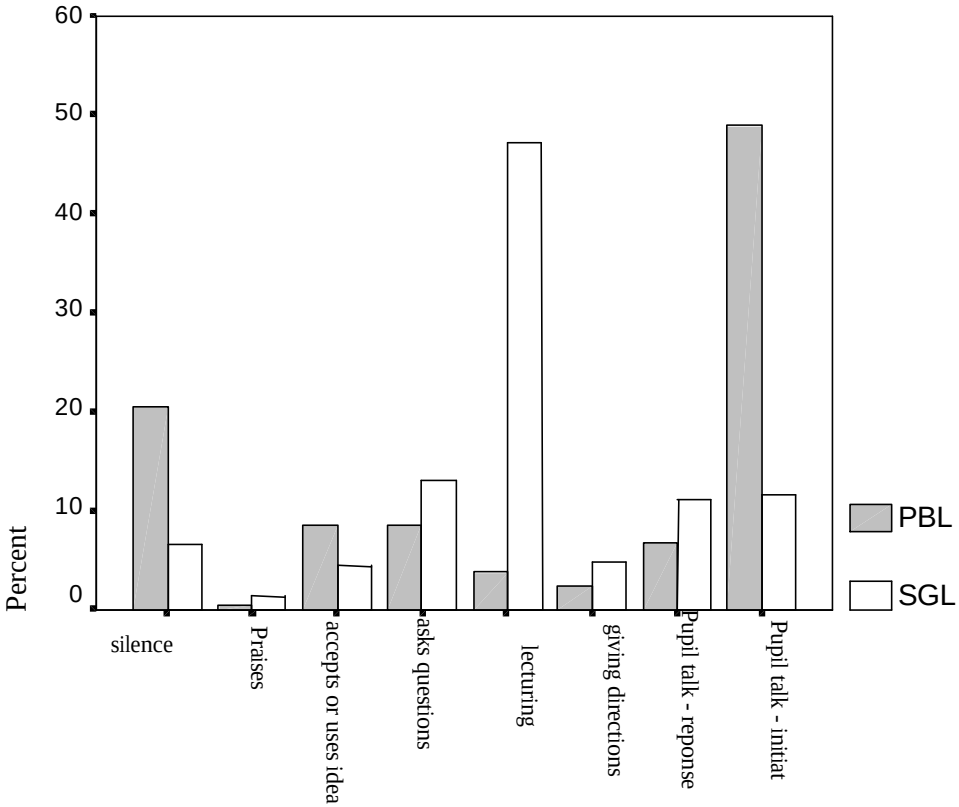


Figure 2.15 Comparing patterns of classroom interaction analysis in the PBL and SGL curricula



The interaction tallies for each teacher in the respective curricula were added together and then divided by the number of teachers in the curriculum to give an estimated average pattern of interaction in each curriculum. Figure 2.15 shows the results of this comparison. This diagram illustrates quite clearly the differences in patterns of classroom interaction between the experimental (PBL) and control (SGL) curricula. In the control (SGL) curriculum classes approximately 65% of the interaction was taken up by teacher centred interaction whereas in the experimental (PBL) curriculum classes this figure was approximately 20%. In the experimental (PBL) curriculum classes, approximately 50% of the interactions were student initiated compared with approximately 10% in the control (SGL) curriculum classes. Also important is the greater proportion of time spent in silence in the experimental (PBL) curriculum classes, as this suggests that silence in which thinking may occur is tolerated by the teacher (Black & Wiliam, 1998). In these respects the teaching approach in the experimental (PBL) curriculum appears, to use Flanders' vocabulary, more 'student centred'.

Summary of similarities and differences between the experimental (PBL) and control (SGL) Curricula

The similarities and differences between the curricula in terms of the key features of Problem Based Learning that have been highlighted in the preceding analysis and discussion are given in box 2.13 below. The table is based on the curricula as practised as opposed to as designed or intended. The wording used to describe the similarities and differences makes use of terminology employed in a number of Problem Based Learning models (Charlin et al, 1998; Savin-Baden, 2000). The Diagrammatic schema is based on that used by Harden & Davies (1998) to differentiate between educational strategies on a continuum. This is based on the argument that concepts and principles are built out of examples and therefore subject matter of programmes can be classified by the way rules and examples are used. Using this schema the

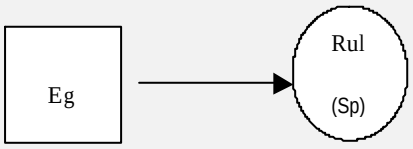
Experimental (PBL) curriculum would fall into the category 'Problem Centred Discovery Learning' and the control (SGL) curriculum into the category 'Problem Oriented Learning'.

What is important from the point of view of the question in this study is that the experimental (PBL) curriculum is clearly different from the control (SGL) curriculum. However, for the purposes of comparison with other and future studies it is useful to review where this particular example of a Problem Based Learning Curriculum fits in terms of the key features of the approach. The Harden & Davies (1998) continuum differentiates between curriculum based solely on conceptions and principles but even on this basis it not easy to characterise the Experimental (PBL) curriculum used in this study as 'a' particular type. Savin Baden (2000) identifies six different elements in her framework of Problem Based Learning on the basis of which she identifies five different models of Problem Based Learning. The experimental (PBL) curriculum does not fall neatly into anyone of these styles as the curriculum is inconsistent across the six elements in her framework. It is suggested that the Problem Based Learning curriculum in this study falls somewhere between Savin Baden's model III 'Problem Based Learning for interdisciplinary understanding' and model IV 'Problem Based Learning for Transdisciplinary Learning'.

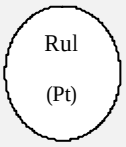
The experimental (PBL) curriculum may not have been 'optimal' in terms of its design and/or implementation on a number of the key features Problem Based Learning outlined in part I of the report. Firstly the students only engaged in self and peer evaluation to a limited extent. Secondly the Problem Based Learning Groups did not all reach a stage where they were functioning well as a group. Thirdly, the student's feedback tended to be limited to 'mini lectures'. Fourthly, students did not always engage in elaboration. Finally, the programme assessments only gave limited attention to the assessment of meta-cognitive and interpersonal skills. The possible implications of this are discussed in part III of the report.

<table>
<tr><th colspan="3">Box 2.13: Similarities and differences between the experimental (PBL) and control (SGL) curricula</th></tr>
<tr><th></th><th>'Experimental (PBL)curriculum</th><th>Control (SGL) Curriculum</th></tr>
<tr><td>Conception of knowledge, teaching and learning</td><td colspan="2">Similarities
View of learning as the outcome focused acquisition of knowledge and skills for the workplace
Assessment constructed as the opportunity to demonstrate an integrated understanding of skills, personal and propositional knowledge across disciplines

Differences</td></tr>
<tr><td></td><td>• Focus on examining and testing out of given knowledge and frameworks
• Students viewed as independent thinkers who take up a critical stance towards learning
• Teachers viewed as a coordinator of knowledge and skill acquisition across boundaries
• Focus on learning as an active reflective process



'Problem Centred Discovery Learning'</td><td>• Focus on Practical and Performative Knowledge
• Students viewed as receivers of knowledge who acquire and understand propositional knowledge through absorption of 'facts'
• Teacher viewed as guide to best practice and understanding of propositional knowledge



'Problem Oriented Learning'.</td></tr>
<tr><td>Classroom activity and interaction</td><td colspan="2">Similarities
• Use of small groups

Differences</td></tr>
<tr><td></td><td>• The Goals/ activities determined by students within curriculum framework
• Includes focus on cognitive skills
• Much student- student interaction
• Some peer regulation
• Some self-regulation
• Little tutor participation</td><td>• Goals/activities determined by instructor
• Emphasis on propositional knowledge
• Little student-student interaction, much instructor - student interaction
• No peer regulation, Instructor corrects, criticises or reject erroneous or irrelevant student contributions
• Little self regulation
• Much instructor participation</td></tr>
</table>

Part II of the report presented an analysis of the experimental (PBL) and control (SGL) curricula using the framework of 'the inner teaching-learning environment'. This process evaluation analysed the key principles in the practice of Problem Based Learning and their operationalisation in the curriculum within the institutional and disciplinary contexts in which the study took place. Part III of the report presents the results of the outcome evaluation and discusses their interpretation and the conclusions which may be drawn from the study findings.

PART III

Results discussion and conclusions

Part III Contents

	Page
Progress through the experiment sample size and power at the completion of the programme	**151**
Results: Learner satisfaction	**154**
Course Evaluation Questionnaire (CEQ)	154
Learner satisfaction - Nominal Group Technique	157
Learner satisfaction withdrawal rates and Exit Interviews	160
Learner Satisfaction – Follow-up questionnaires	162
Results - Skills Personal and Propositional Knowledge	**166**
Assignment scores	166
Managers assessment	170
Follow-up questionnaire Student self rating of the performance in practice	174
Follow-up questionnaire Student self –assessment of programme impact	176
Group work video assessment	180
Results self –directed learning -	**183**
ASSIST	183
Follow-up questionnaire - The Self Directed Learning Readiness Scale (SDLRS)	186
Results teacher workload	**190**
Discussion	**192**
Limitations of the study	192
Threats to internal validity - differences in baseline characteristics	192
Other threats to internal validity	195
Threats to construct validity -	
Confounding levels of construct - reliability of intervention	196
Novelty effects	197
measurement issues	198
Threats to statistical conclusion validity	199
Issues in interpreting the results of the study	199
Deriving conclusions	201
Conclusions	**206**
Conclusions about the effects of Problem Based Learning in this study	206
What conclusions from this study can be generalised?	208
Comparison with other similar studies	210
Why did the Problem Based Learning curriculum used in this study not lead to the predicted results ?	214
A problem with Problem Based Learning ?	220
Future developments in research on Problem Based Learning	223

Figures **Page**

Figure 3.1: Flow diagram – Progress through the phases of the randomised field trial 151
Figure 3.2: Programme completion rates in the two curricular 160
Figure 3.3: Scatterplot of students' ratings on the two satisfaction scales in the follow-up questionnaire 164
Figure 3.4: Scatterplot: Inter-rater reliability care plan marking 167
Figure 3.5: Scatterplot: Inter-rater reliability literature review marking 167
Figure 3.6: Scatterplot: Inter-rater reliability Action plan marking 168
Figure 3.7: Scatter plot managers rating of performance against students self-rating of performance 175

Boxes

Box 3.1: Formula for calculating 'effective' sample size (N) with unequal groups 152
Box 3.2: Course Experience Questionnaire Scales and items 154
Box 3.3: Follow- up survey satisfaction scale items 163
Box 3.4: items in the managers' performance rating scale 171
Box 3.5: items in the students self-rating of performance scale in the follow –up questionnaire 174
Box 3.6: Follow-up questionnaire self –rating of programme impact scale 178
Box 3.7: Group performance criteria used in video assessment 181
Box 3.8: Short ASSIST scale items 183
Box 3.9: ASSIST Scale Internal consistency and correlation between subscales 184
Box 3.10: Formula used for calculating change effect size for the ASSIST instrument 185
Box 3.11: Scales and items in Self directed Learning Readiness Instrument 186

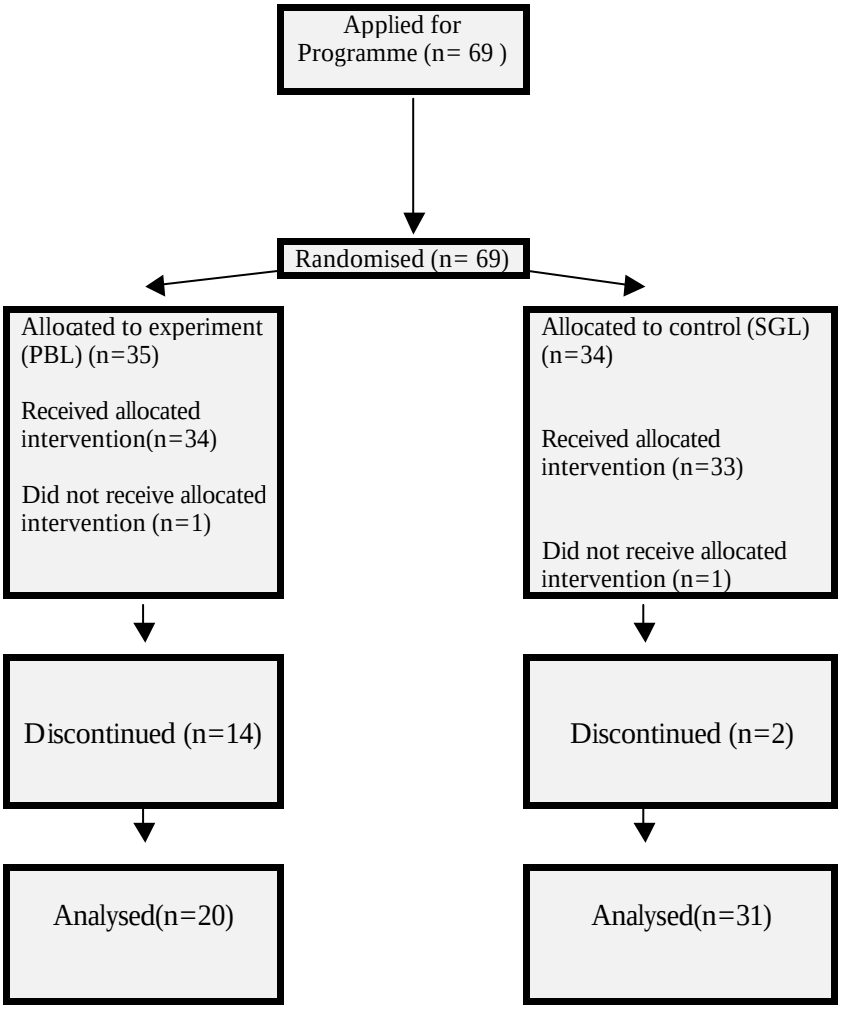
Tables **Page**

Table 3.1 CEQ response scales 155
Table 3.2: Comparison of mean Course Evaluation Questionnaire scale scores 155
Table 3.3: Course Evaluation Questionnaire Scale internal consistency 156
Table 3.4: Sensitivity analysis CEQ scale scores with 'missing' data imputed 156
Table 3.5: Nominal Group Technique results – Things liked about the programme 158
Table 3.6: Nominal Group Technique results things disliked or difficult about the programme 159
Table 3.7: Mean Scores on Pre-administration of Short ASSIST scales finishers & non-finishers 161
Table 3.8: Follow- up survey student ratings of strengths and weaknesses of their programme 163
Table 3.9: Student follow - up survey. Multi -item satisfaction scale comparison Mann Whitney U test 165
Table 3.10: Student follow-up survey: Satisfaction scale results by scale rating category 165
Table 3.11: Assignment first and second markers 166
Table 3.12: Assignment: Inter rater agreement - Kappa scores 169
Table 3.13: Comparison of standard deviated Z scores, effects sizes and sensitivity analysis for assignment marks 170
Table 3.14: Managers follow –up survey: Number of months managed 'student' 171
Table 3.15: Manager follow-up survey. Position of students nominated line manager 172
Table 3.16 : Manager follow –up survey: Global rating of performance 172
Table 3.17: Mann Whitney U test of comparing Managers Global assessment of performance 172
Table 3.18: Managers rating of students work performance multi-item scale and sensitivity analysis 173
Table 3.19: Follow-up questionnaire, Students self-rating of performance 175
Table 3.20: Student follow-up survey, Activities undertaken that maybe indicators of programme impact 176
Table 3.21: Characteristics of follow- questionnaire sample at start of educational programme. 177
Table 3.22: Student Follow- up questionnaire, transformed impact response scales 178
Table 3.23: Follow-up questionnaire, Students rating of programme impact 179
Table 3.24: Mann Whitney U test results - Follow-up questionnaire students rating of programme impact 180
Table 3.25: Student Follow-up questionnaire – Impact scale score less than zero 180
Table 3.26: Ranking of group performance on video assessment by assessor 181
Table 3.27: Mann Whitney U test results - rating of group performance on video assessment 182
Table 3.28: Mean ASSIST scale scores at T^2 and sensitivity analysis 185
Table 3.29: Comparisons of changes in mean ASSIST scale scores T^1 -T^2 186
Table 3.30 Comparison of means n the Self Directed Learning Readiness and subscales 187
Table 3.31: Student Follow-up questionnaire – Self Directed Learning Readiness Total scale score <150 188
Table 3.32: Correlation's between student scores on scales in the ASSIST and Self Directed Learning Readiness Instruments 189
Table 3.33: Teachers workload mean number of minutes teaching per week 191
Table 3.34: Baseline characteristics and outcome variables tested for association 193
Table 3.35: Statistically significant tests of association between characteristics of the sample and outcomes 194
Table 3.36: Criteria for levels of warrant 202
Table 3.37: Summary, weighting of evidence and warrant for conclusions for category student satisfaction, and classroom teaching time 203
Table 3.38: Summary, weighting of evidence and warrant for conclusions category approaches to learning 204
Table 3.39: Summary and weighting of evidence, skills personal and propositional knowledge category 205

Progress through the experiment sample size and power at the completion of the programme

The sample size calculations used for the study are given in part I of the report. Figure 3.1 below shows how the sample size changed in the experimental (PBL) and control (SGL) curricula through the study. In each arm of the study one student who registered did not attend at the start of the programme. During the programme the withdrawal rates were different for each group and therefore there is some variation in the sample size for each outcome measure.

Figure 3.1: Flow diagram – Progress through the phases of the randomised field trial



Because the final sample size is derived from two unequal groups, the 'effective' sample size was used for estimating the post-hoc power (Altman, 1991). The formula used for this calculation is given in box 3.1 below. Using a nomogram, at a power of 80% an 'effective' sample size of 48 will detect a standard deviated effect size of 0.8 as 'significant' at the level of 0.05. As stated in

part I the intention of the study was to identify effect sizes of d=1.0 or greater. However different audiences may regard different effect sizes as important and as previously stated the approach of the analysis is oriented towards estimation rather than hypothesis testing (Altman, 1991). Therefore all effect sizes are reported along with their respective confidence interval. The confidence interval is useful as it provides an indication of the range within which the true' i.e. population difference may lie.

Box 3.1: Formula for calculating 'effective' sample size (N) with unequal groups (Altman, 1991)

$$N = 4N'k/(1+k)^2$$

$$K = n_1 / n_2$$

$$N' = n_1 + n_2$$

n_1 = final sample size for intervention group and n_2 = final sample size for control group

The strategy and process of data analysis is described in part I of the report. Unless otherwise stated data were normally distributed with homogeneity of variance and were analysed using the independent samples T-Test. The term 'effect' is used to refer to an outcome measure of the study. Effects are often reported in different units making comparison within and between studies difficult. One approach is to report outcomes as standardised effect sizes (d). The effect size is the standardised mean difference between the two groups and thus provides an estimate of the size of any difference. One particular advantage of this approach is that we do not need to be familiar with the scale used by the researcher in order to interpret the result (Coe, 2002b). There are a number of approaches that can be used to calculate effect sizes. The Meta-Stat software used for the analysis of the assignment data uses the control group standard deviation as the denominator (Rudner et al, 2002) therefore this method was used for calculating effect size throughout.

Data may be incomplete for many reasons. The most important data relates to participants who drop out of the study before the end. Where there are many more withdrawals in one group the results of the study will be compromised (Altman, 1991). As indicated below the withdrawals for the experimental (PBL) curriculum were much greater than from the Control (SGL)

curriculum. There are a number of different approaches to sensitivity analysis i.e. analysis that estimates the impact of withdrawals. Unless otherwise indicated the approach used in the sensitivity analysis reported here is to impute scores for the missing data based on the lowest individual mean score for each curricula.

Results: Learner satisfaction

Course Evaluation Questionnaire (CEQ)

The items and scales that comprise the Course Evaluation Questionnaire are given in Box 3.2 below. For the purpose of analysis where scale items were negatively worded the response values were reversed (i.e. five recoded as one). The likert scale used in the Course Evaluation Questionnaire uses response values ranging from one 'very satisfied' to five 'very dissatisfied'. To calculate a total score the original response values were transformed as shown in table 3.1.

Box 3.2: Course Experience Questionnaire Scales and items

Quality of teaching scale

- The teaching staff of this course motivate students to do their best work
- The staff make a real effort to understand difficulties students may be having with their work
- Teaching staff here normally give helpful feedback on how you are going
- Our lecturers are extremely good at explaining things to us
- Teaching staff here work hard to make their subjects interesting to students
- This course really tries to get the best out of all its students
- Staff here show no real interest in what students have to say
- Lecturers here frequently give the impression that they have not got anything to learn from students

Clear Goals and Standards Scale

- Its always easy here to know the standard expected of you
- You usually have a clear idea of where you are going and what's expected of you in this course
- Its often hard to discover what is expected of you in this course
- The aims and objectives of this course are not made very clear
- The staff here make it clear right from the start what they expect from students

Appropriate workload scale

- The workload is too heavy
- It seems to me that the syllabus tries to cover to many topics
- We are generally given enough time to understand the things we have to learn
- There is a lot of pressure on you as a student here
- The sheer volume of work to be got through on this course means that you can't comprehend it all thoroughly

The appropriate assessment scale

- Staff here put a lot of time into commenting on students work
- To do well on this course all you really need is a good memory
- Staff here seem more interested in testing what we have memorized than what we have understood
- Too many staff ask us questions just about facts
- Feedback on student work is provided only in the form of marks and grades
- It would be possible to get round this course just by working hard around exam times
- There is very little choice in this course in the ways you are assessed

The emphasis on independence scale

- There are few opportunities to choose the particular areas you want to study
- The course seems to encourage us to develop our own academic interests as far as possible
- Students have a great deal of choice over how they are going to learn in this course
- Students here are given a lot of choice in the work they have to do
- We often discuss with our lecturers or tutors how we are going to learn in this course

Table 3.1: CEQ response scales

	Very satisfied		Neutral		Very dissatisfied
Standard response scale	1	2	3	4	5
Transformed score	+100	+50	0	-50	-100

Table 3.2 below gives the results of the analyses comparing the mean scores on each scale between the intervention and control groups. The results in this table are based on the mean Course Evaluation Questionnaire scale score for each individual student. Only students that completed all three Course Evaluation Questionnaires are included. The results on each scale indicate a higher level of satisfaction amongst students in the control (SGL) curriculum. On three of the five scales the difference in effect size is 'important' and statistically significant.

Table 3.2: Comparison of mean Course Evaluation Questionnaire scale scores

Scale	Sample size Exp Control Start /effect	1- Exp. group (s.dev)	2- Cont. group (s.dev)	1-2	Effect size *d*	95% C.I	P
Workload		-24 (115.3)	101 (119.8)	-125	-1.0	-0.4 to –1.6	0.00
Teaching		289 (163)	568 (134.9)	-279	-2.0	-1.4 to -2.8	0.00
Goals & standards	34/21 33/29	39 (128.6)	229 (98.2)	-190	-1.9	-1.3 to -2.6.	0.00
Independence		142 (141.5)	161 (101.9)	-19	-0.2	-0.7 to 0.4	0.57
Assessment		254 (123.5)	308 (125)	-54	-0.4	-1.0 to 0.6	0.13

The reliability of the Course Evaluation Questionnaire (CEQ) has been established in many studies over a number of years (Long & Johnstone 1997). However, where the intention is to compare mean scores on the subscales it important to ascertain the degree of internal consistency for each scale within the particular study sample. Measures of internal consistency estimate how consistently individuals respond to the items within a scale. The internal consistency of the measurement scales were assessed using SPSS to calculate Cronbach's Alphas (α) for each scale. The widely accepted social science cut-off is that alpha should be .70 or higher for a set of items to be considered consistent (Bryman & Cramer, 1995). On both the 'Goals and standards' and 'Independence' scales the control (SGL) curriculum Alpha scores lack consistency and in the 'Assessment' scale both curricula alpha scores indicate poor consistency. It is interesting to note

that on the two scales where the scores of both curricula indicate consistency, 'Workload' and 'Teaching', the standardised effect sizes are considerably larger.

Table 3.3: Course Evaluation Questionnaire Scale internal consistency

Scale	Group	Cronbach's Alpha
Workload	PBL	0.63
	SGL	0.73
Goals and standards	PBL	0.75
	SGL	0.46
Independence	PBL	0.63
	SGL	0.57
Teaching	PBL	0.81
	SGL	0.83
Assessment	PBL	0.43
	SGL	0.37

Sensitivity analysis for Course Evaluation Questionnaire scales used the lowest mean score on each scale from the experimental (PBL) and control (SGL) curricula respectively to substitute for missing data i.e. for those students that did not complete any Course Evaluation Questionnaires. For those students that completed less than three Course Evaluation Questionnaires the average score for those completed was used. It seems reasonable to assume that those students that did not complete the programme were 'less satisfied' than those that did. Sensitivity analysis has the effect of increasing the difference in scores between the two curricula. In the case of the scales 'appropriate assessment' and 'emphasis on independence' the effect sizes move toward a level that has been accepted as 'important' in this study and the difference in the mean scores between the curricula become statistically significant.

Table 3.4: Sensitivity analysis Course Evaluation Questionnaire scale scores with 'missing' data imputed

Scale	Sample size Exp Start / Control effect	1- Exp. group (s.dev)	2- Cont. group (s.dev)	1-2	Effect size *d*	95% C.I	P
Workload	34/33 33/33	-78 (166.9)	77 (131.6)	-155	-1.2	-0.7 to -1.7	0.00
Teaching	34/33 33/33	216 (181.7)	557 (150.4)	-341	-2.3	-1.6 to -2.9	0.00
Goals and standards	34/33 33/33	-54 (175.5)	221 (108.3	-275	-2.5	-1.9 to -3.2	0.00
Independence	34/32 33/33	66 (184.8)	151 (106.7)	-85	-0.8	-0.3 to -1.3	0.02
Assessment	34/33 33/33	188 (148.9)	280 (140.5)	-92	-0.7	-0.2 to -1.2	0.12

At the end of the Course Evaluation Questionnaire, there was a space where students could make any additional comments about the programme if they wished. The Course Evaluation Questionnaires were typically handed out in the classroom for immediate completion, which probably reduced the likelihood of students completing this section. Students were consistent in their pattern of commenting they tended to either always write comments or never write comments. Given this tendency, caution is required in interpreting these results as they are heavily influenced by a small number of individuals. In the experimental (PBL) curriculum 43% (26/61) of returned Course Evaluation Questionnaires contained comments and in the control (SGL) curriculum 35% (33/93). Comments usually comprised of one or two statements about the programme. Each individual comment could contain both positive and negative statements. There were marked differences between the statements made by students in the experimental (PBL) and control (SGL) curricula. In the experimental (PBL) curriculum there were 14 statements about the heavy workload. In addition to this there were 13 negative statements about the programme, the majority of which were about the lack clarity about what was expected of them and difficulties with Problem Based Learning. There were also 11 statements about how enjoyable the programme was. In the control (SGL) curricula there were 45 positive statements about the programme the majority of which were about the quality of the teaching/teacher and the interest and relevance of the course. The few negative statements related mainly to a specific incident that occurred in one of the control (SGL) curriculum groups.

Results: Learner satisfaction - Nominal Group Technique

The Nominal Group Technique was carried out for each group on the last scheduled 'teaching' day. Students had completed their assignments and the programme was essentially 'over'. In most of the groups all the students who had not withdrawn from the programme attended. There was no independent verification of the veracity with which each group followed the

prescribed process. The groups wrote up their lists on a flip chart. All of the items listed by each group were aggregated for each curriculum.

Table 3.5: Nominal Group Technique results – Things liked about the programme

Experimental PBL Curriculum	Control (SGL) Curriculum
• Working together as a group – support friendship, learning from each other • Advanced Life Support Training (outside PBL framework) • Getting motivation to study • Presentations – gained confidence • Discussion/ brainstorming • ECG workshop (outside PBL framework) (yr1) • Learning about PBL • Identifying own learning needs • Improving IT skills • Visiting clinical areas • Some topics	• Specialist input/ study days • Advanced Life Support Workshop • Hearing from other group members • Programme leader (teacher) • Doing presentation • Programme was Challenging • Programme well organised and structured • Small group size

Table 3.5 lists all the items described as 'liked' by the groups in both curricula in descending order of frequency. The lists of items put forward by the control (SGL) curricular groups usually included more than one specific 'specialist' teaching session. Given the intention of this exercise was to gain insight into students likes and dislikes in terms of course organisation, cultures and learning activities these were amalgamated into a generic category 'specialist input'. Hence, fewer items appear in the responses from students in the control (SGL) curriculum. The top three items on the control (SGL) curriculum groups list were mentioned by all of the groups and the last three only one. Another noteworthy difference is that three out of the four groups in the control (SGL) curriculum included 'the teacher' on their list of likes whereas none of the groups in the experimental (PBL) curriculum did so. Two groups in the experimental (PBL) curriculum stated that they liked learning about Problem Based Learning.

Table 3.6: Nominal Group Technique results things disliked or difficult about the programme

Experimental (PBL) curriculum	Control (SGL) curriculum
• Doing PBL	• Aspects of assignments / Assessment
• Workload	• Location of teaching sessions / travel
• Aspects of assignments	• Working & studying (workload)
• Group dysfunction	• University administration
• Lack of input from Teacher	• Research
• Doing presentations	• Timetable - Overlapping modules
• Being the chair	• Classrooms
• Evaluating each PBL session	• Certain sessions / topics not useful
• Lack of choice of teaching method	• Having to do presentation
• Presentations	• Other class members attitudes
• Difficulty in knowing what is wanted	• Not enough emphasis on practical skills

Table 3.6 shows the items listed as 'disliked' or 'found difficult' for each curriculum. In both curricular the students felt that the workload was difficult to manage whilst working as well and that they disliked the programme assignments. One group in each of the curricula also listed having to do presentations as a dislike. What is strikingly different about the lists is how the dislikes of students in the control (SGL) curriculum appear to be concerned with administrative / organisational aspects of the programme whereas the dislikes listed by students in the experimental (PBL) curriculum are largely related to the 'teaching approach' used. Problem Based Learning was at the top of the list of dislikes in each group in the experimental (PBL) curriculum even though two of the groups also included Problem Based Learning on their 'like' list.

After the Nominal Group Technique exercise was completed the groups were asked to elaborate / explain their lists. Some groups were more interested in participating in this discussion than others. The groups in the control (SGL) curriculum in particular did not wish to engage in discussion and only responded to direct questions. Perhaps because they were sensitised to the 'experimental' status of Problem Based Learning, the groups in the experimental (PBL) curriculum were more forthcoming. The impression of students' views about Problem Based Learning that emerged from the discussion might be viewed as more positive than the Nominal Group Technique results. In general, students appeared to be slightly confused about their view of Problem Based Learning saying that they recognised the value of Problem Based Learning even if they did not enjoy it as one student put it

" I really did not like doing it by Problem Based Learning, but looking back I am glad I did it that way as I can see that it is the most appropriate way of learning for the future"

(Student NGT discussion PBL group 3)

As part of the discussion the students in the experimental (PBL) curriculum groups were asked whether if in future they would undertake a course if the teaching & learning strategy used was Problem Based Learning. About a third of the students said no, a third said yes and for the remaining third the answer was a pragmatic 'it depends'.

Results - Learner satisfaction - Withdrawal rates and Exit Interviews

Figure 3.2: Programme completion rates in the two curricular

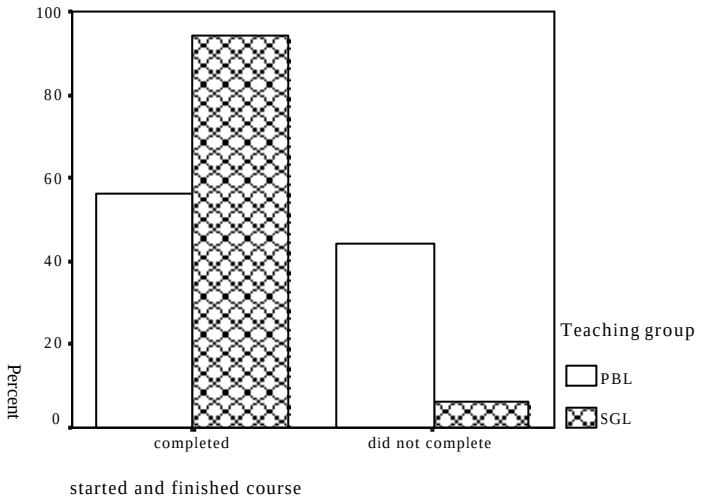


Figure 3.2 illustrates the difference in programme completion rates between the two curricula. 31 (94%) of students in the control (SGL) curriculum completed the programme compared to 20 (59%) of students in the experimental (PBL) curriculum (Pearsons χ^2 = 11.36 p= 0.0005). Another way of presenting this is that students in the experimental (PBL) curriculum were 36% (95% C.I 18-55) more likely to withdraw than students in the control (SGL) curriculum. For every three students who followed the PBL curriculum one would not complete it who otherwise would have done so had they been in the control (SGL) curriculum (95% C.I 2 to 6). By converting the results to an odds ratio (odds ratio for not completing the programme = 10.8) an effect size of -0.9 (95% C.I -1.4 to -0.4) can be calculated (Hedges, 2003).

The analysis above is based on all the students who started their respective programmes. One student in the experimental (PBL) curriculum withdrew after attending two teaching sessions and one student in the control (SGL) curriculum withdrew after attending one teaching session. Both of these students gave their reason for withdrawal as being unable to manage the workload. The second student who withdrew from the control (SGL) curriculum did so after failing the first assignment at the end of module one. In the experimental (PBL) curriculum two students withdrew in the middle of the first module, nine at the end of the first module and one in the middle of the third module.

Male students were more likely to withdraw from the programme (Difference = 48% 95%C.I 18% to 78% $P<0.01$). Those students who worked on non-general wards were also more likely to withdraw from the programme (38% v 16% Pearsons χ^2 $P=0.05$). The students who withdrew from the programme were similar to those who completed the programme on all the other variables assessed in the pre-questionnaire. The students who withdrew were also likely to have more 'undesirable' approaches to learning as assessed by the short ASSIST instrument administered at the beginning of the programme. Table 3.7 shows the mean scores on the ASSIST scales for those who completed the programme and those that did not. The confidence intervals for the effect sizes exclude zero in each case indicating that the difference is statistically significant.

Table 3.7: Mean Scores on Pre-administration of Short ASSIST scales finishers & non-finishers

Scale	Sample size Completed/ withdrawn		1.Completed (s.dev)	2. Not completed (s.dev)	1-2	Effect size *d* (95% C.I)
Surface Scale	50	16	17 (4.7)	22 (3.3)	-5	-1.5 (-2.1 to -0.9)*
Deep Scale	50	16	25 (2.9)	22 (2.3)	3	1.3 (0.7 to 1.9)
Strategic Scale	50	16	23 (4.5)	19 (3.6)	4	1.1 (0.5 to 1.7)

*Higher scores on this scale are = worse outcome hence value for column 1 - column 2 is given negative value.

Telephone exit interviews were conducted with all except one student. This student was not interviewed, as the fact the student had withdrawn was not made known to the Principal Investigator until after the programme was completed. Attempts to contact this student at this

point were unsuccessful. Six of the students who withdrew from the experimental (PBL) curriculum stated that they did so for personal/ work related reasons which meant that they 'had' to leave the programme e.g. they changed jobs. The data presented below relates only to those students who cited Problem Based Learning as being their main reason for withdrawing (n=7). However even where this is the case it should be noted that in four cases external factors required them to withdraw (leaving job, long-term sick leave). For three out of the fourteen students there would appear to be no other grounds for withdrawal apart from their stated dislike/ problems with Problem Based Learning. The seven students felt that organisation of the experimental (PBL) curriculum was poor particularly at the beginning of the programme. The overall feeling of these students was that the content of the experimental (PBL) curriculum was unclear. Five out of the seven students used the term 'have not learnt anything'. All reported that they found Problem Based Learning very difficult, stating that their experience was one of continuous struggle with little progress. In five out of the seven cases the workload associated with the programme was heavier than expected, in their view because Problem Based Learning required too much independent study. The overall view of these students was either that there was no teaching or that the teaching quality was poor. Some felt that this was because the teachers lacked the skills for Problem Based Learning and/or had insufficient subject knowledge. Whilst the majority of students who withdrew from the experimental (PBL) curriculum had experienced a change in their personal circumstances it appears that they all had been dissatisfied in one way or another with their experience in the experimental (PBL) curriculum.

Learner Satisfaction – Follow-up questionnaires

The student follow-up survey included a scale measuring students satisfaction with the programme (see box 3.3 for the scale items). Each item was rated on a five point likert scale with 1 being a 'definite weakness' and 5 being a 'definite strength'. Analysis of the scale involved computing a total scale score for each student (scale maximum = 70) and comparing the difference in the mean scores for each of the curricula. The intra-item correlation coefficients

were satisfactory for both the experimental (PBL) (∝=0.93) and Control (SGL) (∝=0.82) curricula. Students were also asked to summarise how they felt about the way they had learnt on their programme using a numerical scale where 1 = 'All students should not learn the way I did', to 10 = 'All students should learn the way I did'. The questionnaire was mailed out to students between 4-8 months after the completion of the programme. The sample size figures here are therefore based only on those students who completed the programme.

Box 3.3: Follow- up survey satisfaction scale items

To what extent you feel the following aspects of the programme are strengths or weaknesses:

- The emphasis on self directed learning
- The amount of independent study required
- The flexibility of the programme
- The learning resources made available by the teacher
- The teaching and learning strategy used
- The organisation of the programme
- The study workload required
- The anxiety level created
- The methods of assessment used
- The performance of the class teacher
- The opportunities given for self reflection and evaluation
- The support given to learners
- The relationship between the students
- The relationship between the students and the teacher

Table 3.8: Follow- up survey student ratings of strengths and weaknesses of their programme

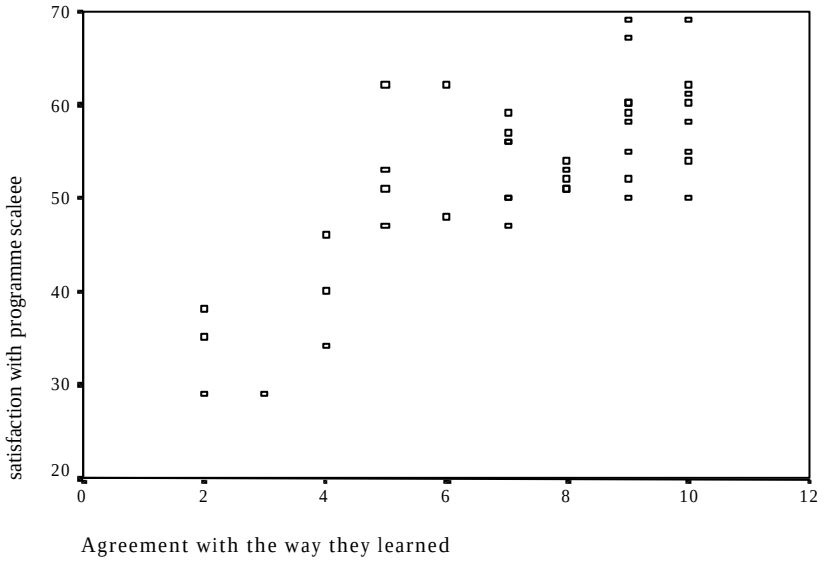
Scale	Sample size Exp Start / Control effect	1- Exp. group (s.dev)	2- Cont. group (s.dev)	1-2	Effect size *d*	95% C.I	P
Multiple item 'Satisfaction'	20/18 31/25	48 (11.6)	56 (6.0)	-8	-1.3	-0.6 to -1.9	0.01*
Multiple item satisfaction (M)	20/20 31/31	47 (11.4)	55 (5.9)	-8	-1.4	-0.7 to -2.0	0.01*
'Taught this way- global scale'	20/19 31/24	6 (2.7)	8 (1.9)	-2	-1.0	-0.4 to -1.7	0.03*
'Taught this way- global scale' (M)	20/20 31/31	6 (2.9)	8 (1.8)	-2	-1.0	-0.5 to -1.7	0.003*

*Homogeneity of variance not assumed

Table 3.8 shows the results of the comparison in mean scores between the two groups on both the multiple item 'satisfaction' scale and the single global rating. The effect sizes on both the multiple item scale and the single global rating favour the control (SGL) curriculum. On both scales the effect sizes are important and the confidence interval excludes zero. Given the distance between starting the programme and the follow-up survey, sensitivity analysis based on all those who started the experiment does not seem to be appropriate in this case. Sensitivity analysis based on non-respondents to the follow-up survey was carried out using lowest score from each curriculum as imputed values. However, this did not lead to results that were particularly different.

If both the multiple item satisfaction scale and the global 'taught this way' scale were tapping into the same construct a positive correlation between the students ratings on the two scales would be expected. The visual correlation evident in the scatterplot in figure 3.3 is confirmed by the correlation analysis (r=0.76 P=0.00).

Figure 3.3: Scatterplot of students' ratings on satisfaction scales in the follow-up questionnaire



It could be argued that the five point response used in the multi-item satisfaction scale represents ordinal rather than interval level data and therefore a non parametric test should be used to

compare the results. The results of the comparison of scores between the experimental (PBL) and control (SGL) curricula using the Mann Whitney U test is given in table 3.9

Table 3.9: Student follow - up survey. Multi -item satisfaction scale comparison Mann Whitney U test

	Mean Rank	Sum of Ranks	Z	P
PBL (n=18)	16.83	303	-2.29	0.02
SGL (n=25)	25.72	643		

The Mann Whitney U value is the number of pairs where the experimental score is lower that the control score. When this figure is expressed as a proportion of all possible pairs this is equivalent to the estimated probability that any new observation in the experimental group will be lower than any new observation in the control group (Altman, 1991). In this case the Mann Whitney U score is 132 and the total number of pairs possible is 450 meaning that any new observation in the experimental (PBL) curriculum has an estimated 29% chance of being lower than any new observation in the control (SGL) curriculum.

An alternative analysis is to compare the two programmes in terms of the original categories used on the ratings scale. The results of such an analysis are shown in table 3.10. In this comparison the difference between the two curricula remains statistically significant (Pearsons χ^2 =10.32 P=0.003). However this comparison makes clear that none of the students in either curriculum rated their experience below average.

Table 3.10: Student follow-up survey: Satisfaction scale results by scale rating category

	Average (%)	Strong (%)	Definitely Strong (%)	Total
PBL (n=18)	6 (33)	8 (44)	4 (22)	18 (100)
SGL (n=25)		13 (52)	12 (48)	25 (100)

Results - Skills Personal and Propositional Knowledge

Assignment scores

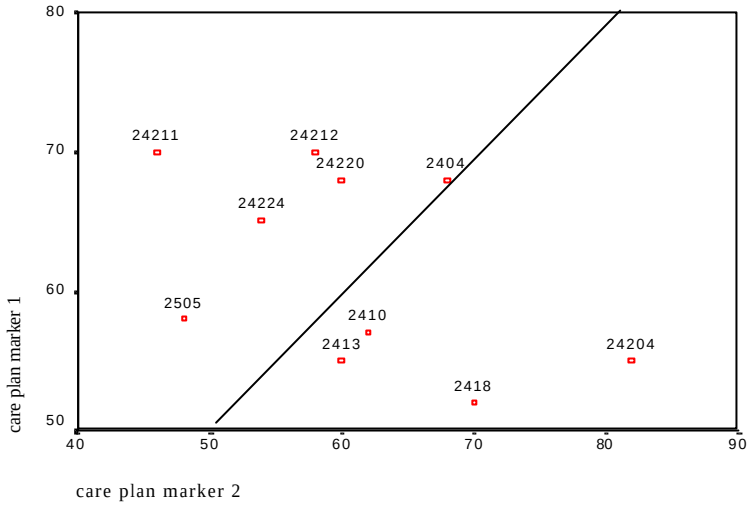
For the research score the three assignments were marked independently by external 'experts'. Each assignment was marked by a different marker to increase the reliability of the overall final assessment mark. The same markers were used for both the first and second cohorts and marked the same paper in both years. The primary purpose of the collecting the data on student assignment scores is to measure whether there are any differences in attainment between the experimental (PBL) and control (SGL) curricula. Group scores for each individual assignment were compared but the primary outcome measure in this section is the average score for the three marks combined. A combined measure of three assessments of performance should provide a more valid and reliable indicator of performance than any single assessment.

The method used to combine scores on the three assignments depends on how similar or different the score given by each marker is for each different assignment. For example, it would not be appropriate to simply use a mean score for the three assignments if the markers were in fact applying the marking guidelines differently. The consistency of the marking between examiners was investigated by measuring inter-rater agreement (or reliability). In addition to marking all the scripts from one assignment each examiner marked five scripts from one of the other assignments in each year (see table 3.11 below). This was the maximum number of scripts possible under the terms of the agreement with the individual markers.

Table 3.11: Assignment first and second markers

Marker	Main assignment	Sample assignment
1	Literature review	Care plan
2	Care Plan	Action Plan
3	Action plan	Literature review

Figure 3.4: Scatterplot: Inter-rater reliability care plan marking

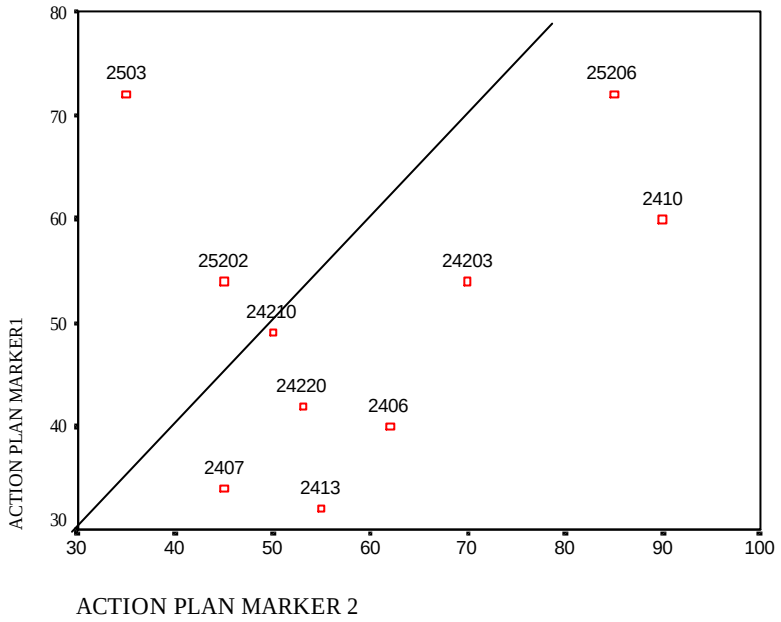


Figures 3.4 to 3.6 are scatterplots that illustrate the inter-rater reliability for each assignment. The marks given by each examiner for the same assignment by the same student are plotted against each other. The point where the two scores intersect is marked by a case number. The line superimposed on each plot approximates perfect agreement. The visual presentation suggests considerable differences between markers in the grading of the same scripts.

Figure 3.5: Scatterplot: Inter-rater reliability literature review marking



Figure 3.6: Scatterplot: Inter-rater reliability Action Plan marking



The level of agreement between markers can be estimated statistically using kappa (*k*). This approach is preferable to the use of Pearson's correlation coefficient as the latter is a test of linear association rather than agreement and is susceptible to a high degree of correlation even when actual agreement is poor (Altman, 1991). Kappa can be interpreted as the 'chance corrected proportional agreement'. It is a measure of agreement that takes into account the degree of achievement that would be expected by chance alone. The calculation is based on the fact that the expected frequency in a cell of a frequency table (under the null hypothesis of no association) is the product of the total of the relevant column and the total of the relevant row divided by the grand total. This gives the agreement expected just by chance. The maximum agreement is 1.00. Kappa thus calculates the agreement between markers as a proportion of the possible scope for doing better than chance (Altman, 1991). The approach required the interval level data from the assessment marks to be converted into categorical data. In the analysis the categories 0-20,21-40, 41-60,61-80, and 81-100 were used.

As the data in table 3.12 illustrate the kappa scores suggest that the level of agreement between markers is worse than would be expected by chance (i.e.0). There is no definite interpretation of kappa but a kappa score of less than 0.2 is regarded as poor (Altman, 1991). The sample size for

each marker is small (n=10) and the unweighted kappa score a crude measure of agreement. However, levels of agreement as low as that found in table 3.12 suggest that a mean score for the three assignments based on raw data would not a reliable indicator of each student's performance.

Table 3.12: Assignment: Inter rater agreement - Kappa scores

Assignment	Kappa	95% Confidence interval
Literature review	-0.07	-0.14 to –0.001
Care plan	-0.27	-0.43 to 0.11
Action Plan	-0.03	-0.16 to 0.1

Where variation between markers appears to be indicated, one approach is to transform the raw scores into standard normal deviated scores to a mean of zero and a standard deviation of 1. Standard normal deviated or 'Z' scores assume that that the scores from each marker form a normal distribution for that assignment and thus a standard deviation from the assignment mean can be calculated for each case. Thus while assignment scores may vary, deviations from the mean remain a constant and comparable feature between assignments (Kutnic et al, 1997). When the standard deviated score is used the mean experimental (PBL) curriculum score minus the mean control (SGL) curriculum score is equivalent to the standardised effect size *d* (Coe, 2002a).

The mean standard deviated Z scores and effect sizes for the difference between the two curricula are given for each assignment in table 3.13 below. Comparison of group scores for each assignment was carried out using a one way ANOVA to test the null hypothesis that there was no difference between mean scores in the two groups (Kinnear & Gray, 2000). The first four effects report the results of the analysis only for the assignments that were submitted i.e. excluding missing data. For each of the individual assignments and for the mean score of the three assignments, the effect size favours the experimental (PBL) curriculum. However, the differences are not statistically significant and the 95% confidence intervals do not exclude an effect in favour of the control (SGL) curriculum.

Table 3.13: Comparison of mean standard deviated Z scores, effects sizes and sensitivity analysis for assignment marks

Effect Standard deviated Z-scores	Sample size Exp Start	Control /effect	1- Int. group (s.dev)	2- Cont. group (s.dev)	1-2 = effect size *d*	95% C.I	F	P
Literature review	34/22	33/32	0.1 (1.22)	-0.07 (0.82)	0.2	-0.4 to 0.7	0.412	0.26
Care plan	34/20	33/31	0.1 (1.23)	-0.08 (0.82)	0.2	-0.4 to 0.8	0.468	0.24
Action plan	32/20	33/31	0.04 (1.24)	-0.03 (0.82)	0.1	-0.5 to 0.7	0.64	0.4
Mean for all assignments	34/20	33/31	0.1 (0.93)	-0.05 (0.58)	0.1	-0.3 to 0.6	0.419	0.4
Sensitivity analysis with imputed 'missing' values								
Literature review			-0.01 (1.1)	0.01 (0.89)	-0.02	-0.5 to 0.5	0.02	0.9
Care plan			-0.04 (1.1)	0.05 (0.88)	-0.1	-0.6 to 0.4	0.162	0.7
Action plan	34/34	33/33	-0.2 (1.12)	0.1 (0.84)	-0.3	-0.8 to 1.3	2.12	0.15
Mean for all assignments			-0.1 (0.88)	0.1 (0.63)	-0.2	-0.5 to 0.2	0.411	0.7

The second four effects reported in table 3.13 are the results of a sensitivity analysis where assignment scores have been imputed for the missing data. In both curricula the lowest individual score for each assignment was so low that it was felt unreasonable to use it as the figure for imputing missing data. Therefore the imputed data for each curriculum is the respective weighted average at the 25th percentile. In the sensitivity analysis the effect sizes for the individual assignments and the mean standard deviated Z score favour the control (SGL) curriculum. However the differences are not statistically significant and the confidence intervals do not exclude an effect in favour of the experimental (PBL) curriculum.

Results - Skills Personal and Propositional Knowledge - Managers assessment

Approximately six months after the completion of the programme the 'current' line managers of the students who completed the programme were sent a postal questionnaire. The former

students were contacted to nominate the person to whom the questionnaire should be sent. The questionnaire asked the manager to rate the former students performance on a single global rating scale and on a multi –item performance scale. Details of the development and pre-testing of the questionnaire are given in part I of the report and the scale items are given in box 3.4 below.

Box 3.4: items in the managers' performance rating scale

Compared to your concept of the ideal team member how would you rate this person's contribution in terms of:

- The quality of their interaction with other members of the health care team
- Their application to work
- Resolving conflicts and disputes
- Reliably performing their job
- Critically evaluating existing practice
- Finding new information for the team
- Managing complex issues
- Influencing other members of the team
- Supporting other members of the team
- Generating good ideas
- The quality of their interaction with patients and their families
- Solving problems
- Identifying and acting on their own learning needs
- Clinical skills and knowledge
- Taking responsibility
- Showing leadership
- Conduct as a professional

The position of the 'line' manager in their respective institutions and the length of time they had been managing the 'student' for whom they were completing the rating scale are given in tables 3.14 and 3.15 below. One manager of a student in the experimental (PBL) curriculum claimed to have been managing the student for 120 months. This was clearly an outlier and distorted the mean for this group and was therefore omitted from the calculation of the mean for the experimental (PBL) group. This case aside the results indicate that the experimental (PBL) and control (SG) groups were reasonably matched in terms of the length of time being managed by the person who responded and the position of the manager in the their organisation.

Table 3.14: Managers follow –up survey: Number of months managed 'student'

Scale	Sample size	Mean (S.Dev)	Range
Length of time managing 'student'	PBL 12	20 (16)	3-54
	SGL 22	21 (15)	1-60

Table 3.15: Manager follow-up survey. Position of students nominated line manager

Position	Experimental (PBL)	Control (SGL)
Clinical/ Directorate manager	5 (36%)	8 (32%)
Ward Manager	9 (64%)	15 (60%)
Team Leader		2 (8%)

Managers were asked to rate the former students' contribution 'compared to their concept of the ideal team member'. Using a five point likert scale ranging from 1 'Unsatisfactory' to 5 'Outstanding'. The results for each curriculum are given in table 3.16. Not all managers completed this scale.

Table 3.16 : Manager follow –up survey: Global rating of performance

Rating	Experimental (PBL)	Control (SGL)
Below Average		1 (5%)
Average	4 (40%)	3 (15%)
Above Average	4 (40%)	11 (55%)
Outstanding	2 (20%)	5 (25%)

Table 3.17 shows the results of a statistical analysis comparing the managers rating of students in the global performance scale between the two curricula. The Mann Whitney U value is the number of pairs where the experimental rating is lower that the control rating. When this figure is expressed as a proportion of all possible pairs this is equivalent to the estimated probability that any new observation in the experimental group will be lower than any new observation in the control group (Altman, 1991). In the case of the managers global rating the Mann Whitney U score is 83 and the total number of pairs possible is 200. Therefore any new observation in the experimental (PBL) curriculum has an estimated 41% chance of being lower than any new observation in the control (SGL) curriculum. However, this difference is not statistically significant.

Table 3.17: Mann Whitney U test of comparing Managers Global assessment of performance

Sample size Start /effect	Mean rank	Sum of ranks-	Mann Whitney U	Z	P
PBL 20/10	13.8	138	83	-0.811	0.24
SGL 31/20	16.35	327			

Table 3.18: Managers rating of students work performance multi-item scale and sensitivity analysis

Scale	Sample size Exp Start / Control effect	1- Int. group (s.dev)	2- Cont. group (s.dev)	1-2	Effect size *d*	95% C.I	P
Multi-item Perform	20/14 31/24	62 (13.3)	65 (10.4)	-3	-0.3	-0.9 to 0.4	0.25
Sensitivity-perform	20/20 31/31	59 (12.1)	63 (9.6)	-4	-0.4	-0.9 to 0.2	0.07

The results of the managers rating of student performance using the multi item performance scale are given in table 3.18. This scale is computed by adding together the ratings given on all the individual items. A larger number of managers completed the items in this scale. The maximum score on the scale is 85 (17 items x 5). The internal consistency of the scale was satisfactory with a Cronbach's Alphas (∝) of 0.97 and 0.96 for the experimental (PBL) and control (SGL) curricula respectively. The range of scores in the experimental (PBL) curriculum was 32-80 and in the control (SGL) curriculum 44-85. The effect size favours the control group but the confidence interval does not exclude zero and the result is not statistically significant.

The second result in table 3.18 is a sensitivity analysis of the possible effect of missing data. The lowest score on the multi-item performance scale in the experimental (PBL) curriculum was 32., and the next 47. It was felt unreasonable to use the lowest score as the figure for imputing missing data. Therefore, the imputed data for each curriculum is the respective weighted average at the 25th percentile. As in the student follow-up questionnaire the effective sample size used was the number of students who completed the programme. The chance corrected proportional agreement between the managers rating on the 'Global' and multi-item performance scales (kappa) = 0.25 (95% C.I 0.18 to 0.3) a level of agreement that is considered 'fair' (Altman, 1991). This suggests that both scales were tapping into the same construct and thus provides more confidence about the reliability and validity of the findings.

Follow-up questionnaire Student self rating of their performance in practice

The follow- up questionnaire asked students' to rate their own performance in practice using a 21-item scale. Details of the development and pre-testing of the scale are given in part I of the report. The items in the scale are listed in box 3.5 below. Students were asked to rate their performance on a five point scale from 1 'strongly disagree' to 5 'strongly agree'. The total score for the scale is obtained by adding the rating for each item. The number of students that completed the programme is used as the 'start' sample in this analysis.

Box 3.5: items in the students self-rating of performance scale in the follow –up questionnaire

- I interact very well with colleagues
- Compared to other nurses in my ward I frequently use published research articles
- In my position appropriate delegation of tasks and responsibilities is more important than getting the job done quickly
- I make short and long term plans about what I have to achieve at work
- My practice (knowledge) is evidence based
- I am confident in my ability to deal with conflict in my workplace
- I encourage other colleagues to participate in discussions
- I am confident in caring for patients from different ethnic or faith backgrounds to my own
- Preventative health care is a very important part of my practice
- People look to me for guidance and help
- I am concerned about what happens to other members of the team
- I am confident in my ability to care for a patient with complex clinical problems
- I trust my colleagues/ team members
- I keep up to date with Government health and nursing policies
- I am able to exert influence with other professionals
- I always assess and plan care to meet patients social and spiritual needs
- I develop good relationships with the patients I care for
- I am confident in my ability to interpret the scientific strength of clinical research evidence
- Listening/talking to patients is a higher priority than managing their clinical care
- My clinical practice is up to date
- I am confident in my ability to work as a member of a team/group

Table 3.19 gives the results of the analysis for the scale total with and without imputed missing values. The internal consistency of the scale was satisfactory with a Cronbach's Alphas ($\propto$) of 0.82 and 0.83 for the experimental (PBL) and control (SGL) curricula respectively The sensitivity analysis (self-perform missing scale) uses the lowest values from the experimental (PBL) and

control (SGL) curricula (78 and 74 respectively) as imputed values. The effect sizes favour the experimental (PBL) curriculum but the confidence interval does not exclude zero.

Table 3.19: Follow-up questionnaire, Students self-rating of performance

Scale	Sample size Exp Control Start /effect	1- Int. group (s.dev)	2- Cont. group (s.dev)	1-2	Effect size *d*	95% C.I	P
Self rating of performance	20/18 31/25	90 (6.6)	87 (6.8)	3	0.4	-0.2 to 1.0	0.07
Self-perform missing	20/20 31/31	88 (7.1)	85 (8.1)	3	0.4	-0.2 to 1.0	0.05

It is interesting to note that the average rating in both curricula are toward the higher end of the scale (maximum scale score 105). Figure 3.7 below is a scatter plot of students own rating of performance and their managers rating of their performance for those students for which both ratings were available (n=34). The visual impression of poor correlation is confirmed by the correlation analysis (Pearson r=-0.07 P=0.35).

Figure 3.7: Scatter plot managers rating of performance against students self-rating of performance



Follow-up questionnaire: Student self –assessment of programme impact

The student follow- up questionnaire contained a number of questions about the students' career development and participation in educational and practice development activities since the completion of the programme. Development and participation in such activities would be expected in nurses even if they had not attended an educational programme. However, differences in rates of development and /or participation between the experimental (PBL) and control (SGL) curriculum may be considered indicative of differential programme impact. The analysis here is testing a series of micro –null hypotheses that participation in an educational programme using Problem Based Learning does not in the short to medium term affect career development or subsequent participation in educational or practice development activities.

Table 3.20: Student follow-up survey, Activities undertaken that maybe indicators of programme impact

Question	Sample size Exp Start	Control /effect	1- Exp. group n (%)	2- Cont. group n (%)	Odds ratio	Effect size *d*	95% C.I	P Pearsons χ^2
Taken up new hobby / interest	20/19	31/26	5 (26)	8 (31)	0.8	-0.3	-0.9 to 0.3	0.3
Participated in formal learning	20/18	31/26	14 (74)	21 (84)	0.83	-0.2	-0.8 to 0.4	0.2
Participated in practice development	20/19	31/24	8 (42)	16 (68)	0.36	-0.6	-1.2 to 0	0.05
Done any teaching	20/19	31/26	5 (26)	8 (31)	0.8	-0.3	-0.9 to 0.3	0.3
Changed ward/ hospital	20/19	31/26	5 (26)	9 (35)	0.67	-0.3	-0.9 to 0.3	0.3
Gained promotion	20/19	31/26	3 (16)	7 (27)	0.51	-0.5	-1.1 to 0.1	0.2

Table 3.20 shows the results of the analysis from the relevant questions. The questions about hobbies and promotion required simple yes/ no responses. The other questions listed in table 3.20 required students to choose from a range of typical activities. For purpose of analysis of the multi response questions the responses were collapsed into dichotomous yes/no categories. For the purpose of this analysis the 'start' sample size for each group given in the table is the number of students in each curricula who completed the programme. The results for each question favour the control (SGL) curriculum. However, with the exception of the question about

participating in practice development the differences between the groups are not statistically significant.

The characteristics of the sample of students who completed the follow-up questionnaires maybe an influencing set of factors on the results to the questions in table 3.20. Table 3.21 below compares the follow-up questionnaire respondents on a range of relevant characteristics reported at the beginning of the programme. Students in the two curricula were similar in terms of the proportion who were married, had children and highest level of formal education.

Table 3.21: Characteristics of follow- questionnaire sample at start of educational programme.

Characteristics	PBL		SGL	
Mean time at current grade	9 months (range 1-34)		12 months *(range 1-48)	
Mean time working in current ward	24* months (range 1-58)		20 months (range 1-48)	
Grade D	4	(21%)	8	(31%)
Grade E	10	(53%)	11	(42%)
Grade F	3	(16%)	6	(23%)
Grade G	2	(10%)	1	(4%)
Undertaken a previous post registration education programme	9	(47%)	18	(69 %)

* One outlying case of 120 months excluded from the analysis

The amount of time working at the current grade and in the same ward are similar for the two groups. If one views the D/E grade and F/G grade as broadly similar in roles then the two curricula would seem to be reasonably evenly balanced. These three sets of characteristics do not suggest that either curriculum group would be expected to be more likely to be involved in practice development or education based on the grade and experience alone. However the experimental (PBL) curriculum group had undertaken less post-registration training which may mean that they were less likely to get involved practice development activities. Against this the fact that they had participated in less post-registration education in the past would suggest that they should have been more likely to have been involved in formal education after the completion of the programme in this study. A greater proportion of the follow-up respondents in the control (SGL) curriculum were in the most junior grade at the start of the programme and thus were more likely to have been promoted in the 18 months between the start of the

programme and completion of the follow-up questionnaire. This suggests that the difference in the proportion of students promoted since completing the programme is in part at least due the different 'start' points of the students in the two curricula.

Students were also asked to rate the impact of the programme on their practice using a 14-item scale. Details of the development and pre-testing of the scale are given in part I of the report. The items in the scale are listed in box 3.6 below

Box 3.6: Follow-up questionnaire self –rating of programme impact scale
Learning that I associate with participating in the programme …….

- Changed my way of thinking about nursing practice
- Changed my way of thinking about using clinical research
- Changed my way of thinking about learning
- Changed my personal attitudes and beliefs.
- Improved my presentation skills
- Improved my ability to deal with conflict
- Improved my ability to work as a member of a team
- Improved my leadership ability
- Improved my nursing care ability
- Improved my ability to manage my own learning and personal development
- Improved my ability to manage information
- Improved my ability to use information technology
- Stimulated me to carry on learning
- Increased my confidence to question clinical policy and practice

Students were asked to rate their agreement with the statements on a five point scale from 1 'strongly disagree' to 5 'strongly agree'. For the purpose of analysis the responses were re-coded (see table 3.22 below). The total score for the scale is obtained by adding the rating for each item.

Table 3.22: Student Follow- up questionnaire, 'impact' response scales

	Strongly disagree		Not sure		Strongly agree
Standard response scale	1	2	3	4	5
Re-coded scale	-100	-50	0	+50	+100

Table 3.23 gives the results of the analysis for the scale total with and without imputed missing values. The number of students that completed the programme is used as the 'start' sample in this analysis. The internal consistency of the scale was satisfactory with Cronbach's Alphas (∝) of 0.96

and 0.91 for the experimental (PBL) and control (SGL) curricula respectively. The sensitivity analysis (impact missing scale) uses the weighted values from the 25th percentile from the experimental (PBL) and control (SGL) curricula (162 and 500 respectively) as imputed values. The total scores were not normally distributed. Both the independent T-test and the effect size calculation are sensitive to the distribution of the data (Altman, 1991; Coe, 2002b). Therefore the effect sizes reported in table 3.23 should be interpreted with caution. Comparison of mean scores of log transformed data using the independent T-test results in a more conservative 'P' value of 0.3.

Table 3.23: Follow-up questionnaire, Students rating of programme impact

Scale	Sample size Exp Start / Control effect	1- Int. group (s.dev)	2- Cont. group (s.dev)	1-2	Effect size *d*	95% C.I	P
Impact	20/18 31/25	617 (630)	798 (393)	-181	-0.5	(-1.0 to 0.2)	0.25
Impact missing	20/20 31/31	571 (612)	740 (371)	-169	-0.5	(-1.0 to 0.1)	0.25

Given that the results on the students rating of programme impact scale were not normally distributed the scores for the experimental (PBL) and control (SGL) curriculum were also compared using the non-parametric Mann Whitney U test. Table 3.24 gives the results of this comparison. The Mann Whitney U value is the number of pairs where the experimental rating is lower that the control rating. When this figure is expressed as a proportion of all possible pairs this is equivalent to the estimated probability that any new observation in the experimental group will be lower than any new observation in the control group (Altman, 1991). In the case of the students rating of programme impact the Mann Whitney U score is 199 and the total number of pairs possible is 450. Therefore any new observation in the experimental (PBL) curriculum has an estimated 44% chance of being lower than any new observation in the control (SGL) curriculum.

Table 3.24: Mann Whitney U test results - Follow-up questionnaire students rating of programme impact

Scale	Sample size Start /effect	Mean rank	Sum of ranks-	Mann Whitney U	Z	P
Follow-up rating of programme impact	PBL 20/18	20.6	370.5	199	-0.629	0.25
	SGL 31/25	23	575.5			

The results of the analysis of the students rating of impact all favour the control (SGL) group however the confidence interval do not exclude zero and the differences are not statistically significant. The 'Total' score of the majority of students in both curricula (PBL 78%, SGL 96%) indicated that they felt that the programme had influenced their 'performance'. The results in table 3.25 below show that the proportion of students whose total score indicated that that they perceived that the programme had no overall impact on their performance is greater in the experimental (PBL) curriculum. The confidence interval for the effect size excludes zero in this case.

Table 3.25: Student Follow-up questionnaire – Impact scale score less than zero

Question	Sample size Exp Control Start /effect	1- Exp. group n (%)	2- Cont. group n (%)	Odds ratio	Effect size *d*	95% C.I	P
Impact scale score <0	20/18 31/25	4 (18)	1 (25)	1.0	-0.8	-1.4 to -0.1	0.09

Group work video assessment

A detailed description of the activities undertaken by the students and methods of assessment used in the group work video assessment are given in part I of the report. Two independent external assessors, who were blind to the allocation status of the groups, assessed the videos independently using a standardised instrument. The criteria against which performance was evaluated are given in box 3.7 below. Both assessors commented that the instrument was difficult to use for the problems/ activities that the groups undertook. In the instrument used the

criteria/statements listed below were matched with a polar opposite statement. Both assessors felt that the polar opposite statement was not useful and so did not use it in their analysis.

Box 3.7: Group performance criteria used in video assessment

- All Members included/involved
- All members attentive in posture
- Widespread interaction between members
- Freedom to express idea thoughts
- Evenly distributed discussion
- Ability to resolve conflict/disagreement
- Critical consideration of all ideas
- Group limits discussion of peripheral issues
- Group demonstrates ability to review progress toward objective
- Group demonstrates coherence of approach
- Group consciously works toward making decision
- Tasks clearly allocated/accepted with group
- Ability to adopt alternative route to solution if one is blocked
- Recognition of appropriate cues in problem presentation
- Hypothesis generation using all available relevant data
- Frequent re-formulation of problem

The assessors took different approaches to grading the groups performance against the criteria. One assessor graded the performance as present or absent for each criteria. The second assessor after finding the above approach insufficiently discriminating graded the groups performance on each of the criteria for each of the three problems that were viewed in the video clips seen by the assessors. This makes direct comparison of the assessor's grading impossible. The ratings given by each individual assessor were computed and the groups placed in rank order for each assessor (see table 3.26 below). There are nine groups because in the second year the control (SGL) curriculum group was split into three groups for the purpose of the video assessment due to the size of the group.

Table 3.26: Ranking of group performance on video assessment by assessor

Rank	Assessor 1 (video group no.-Curriculum)	Assessor 2 (video group no.-Curriculum)
1	1 (Experimental PBL)	1 (Experimental PBL)
2	5 (Control SGL)	2 (Control SGL)
3	2 (Control SGL)	6 (Control SGL)
4	4 & 6 (Control SGL)	7 (Control SGL)
5		8 (Experimental PBL)
6	7 (Control SGL), 8 (Experimental PBL)	4 (Control SGL)
7		5 (Control SGL)
8	9 (Experimental PBL)	9 (Experimental PBL)
9	3 (Experimental PBL)	3 (Experimental PBL)

There were a number of tied groups in assessor one's rankings making direct comparison slightly more complicated. However with the exception of video group five (ranked 2nd and 7th) the rankings appear similar. This impression is confirmed by analysis of the inter-rater agreement resulting in a Kappa score of between 0.4 and 0.5, which is regarded as 'moderate' (Altman, 1991). Given the relative agreement on ranking, the performance ratings given by assessor two were used to compare the two curricula. For each of the criteria a score of between zero and three, based on whether the criteria was observed in none, one, two, or three, of the problems undertaken, was given. A total score was computed for each group by adding together the score for each criterion.

Table 3.27: Mann Whitney U test results - rating of group performance on video assessment

Curriculum (n)	Mean rank	Sum of ranks-	Mann Whitney U	Z	P
PBL (4)	4.25	17	7.0	-0.735	0.23
SGL (5)	5.6	28			

The scores for the experimental (PBL) and control (SGL) curriculum were compared using the non-parametric Mann Whitney U test. Table 3.27 gives the results of this comparison. The Mann Whitney U value is the number of pairs where the experimental rating is lower that the control rating. When this figure is expressed as a proportion of all possible pairs this is equivalent to the estimated probability that any new observation in the experimental group will be lower than any new observation in the control group (Altman, 1991). In this case the Mann Whitney U score is seven and the total number of pairs possible is 20. Therefore any new observation in the experimental (PBL) curriculum has an estimated 35% chance of being lower than any new observation in the control (SGL) curriculum. Not surprisingly given the small sample sizes this difference is not statistically significant. An approximate standard effect size of d= -0.3 was calculated based on the difference in mean scores. However, the 95 % confidence interval ranges from -1.0 to 1.0.

Results self –directed learning -

ASSIST

The items in ASSIST form three scales (see box 3.8 below). The short ASSIST was administered to students at the beginning and end of the programme. The null hypothesis tested is that the experimental (PBL) curriculum does not improve students' approaches to study. Analysis compared the mean scores in each curriculum on the second administration of ASSIST and also the difference between the changes in the mean scores for each curriculum between the first and second administration. In this 'within subjects experiment' each participant acts as her own control making it possible using the within subjects ANOVA to remove the variance associated with individual differences in overall ability (Kinnear & Gray 2000). There were outlying cases on the 'Surface' and 'Strategic' scales in the control (SGL) curriculum on the first administration. There was an outlying case on the 'Deep' scale in the control (SGL) curriculum on the second administration. Log transformation did not improve the distributions and the inclusion of the cases in the analysis did not result in differences of more than 0.4 in the mean score for the respective groups therefore these cases were included in the analysis.

Box 3.8: Short ASSIST scale items

Surface apathetic approach

1. Often I find myself wondering whether the work I am doing here is really worthwhile.
4. I concentrate on learning just those bits of information I have to know to pass.
8. Much of what I'm studying makes little sense: it's like unrelated bits and pieces.
14. Often I feel I'm drowning in the sheer amount of material we're having to cope with.
16. I often worry about whether I'll ever be able to cope with the work properly
18. I often have trouble in making sense of the things I have to remember.

Deep Approach

2. When I'm reading an article or book, I try to find out for myself exactly what the author means.
6. Regularly I find myself thinking about ideas from lectures when I'm doing other things.
10. When I'm working on a new topic, I try to see in my own mind how all the ideas fit together.
12. Often I find myself questioning things I hear in lectures or read in books.
15. Ideas in course books or articles often set me off on long chains of thought of my own.
17. When I read, I examine the details carefully to see how they fit in with what's being said.

Strategic approach

3. I organise my study time carefully to make the best use of it.
5. I look carefully at tutors' comments on course work to see how to get higher marks next time.
7. I'm pretty good at getting down to work whenever I need to.
9. I put a lot of effort into studying because I'm determined to do well.
11. I don't find it at all difficult to motivate myself
13. I manage to find conditions for studying which allow me to get on with my work easily.

The scales are constructed in such a way that the better outcome is a low score on the 'Surface' scale and high scores on the 'Deep' and 'Strategic' scales. The internal consistencies of the scales based on the data in this study are given in box 3.9. The internal consistency scores on the 'Surface' scale are low for both curricula. On the 'Deep' and 'Strategic' scales the internal consistency scores are low for the control (SGL) curricula. Box 3.9 also contains the results of a correlation analysis between the deep and strategic scales. Whilst the expected inverse correlation was found in both groups only in the experimental (PBL) curriculum was the correlation significant at the 0.05 level.

Box 3.9: ASSIST scale internal consistency and correlation between subscales

Scale reliability coefficients (Cronbach's Alpha)

Surface	PBL	r = 0.63	SGL 0.65
Deep	PBL	r =0.72	SGL r =0.53
Strategic	PBL	r= 0.8	SGL r = 0.68

Correlation between Deep and Surface Scales

PBL r= -.452	p=0.001	$100r^2$ = 20%
SGL r= -.156	p=.22	$100r^2$ =2%

Table 3.28 gives the results of the comparison of the mean scores in the two curricula for each scale on the second administration. Because the desirable score on the 'Surface' scale is low the effect size for the difference in mean scores should be interpreted as favouring the control (SGL) curriculum. To avoid confusion these effect sizes have been labelled with a negative sign. However, the 95% Confidence interval for all the effect sizes do not exclude zero and the differences are not statistically significant. The scales with (M) after are the results of the sensitivity analysis. The imputed missing values used were the highest scores on the 'Surface' scale and the lowest scores on the 'Deep' and 'Strategic' scales. When the missing values are included the effect size for the differences in the mean scores between the two curricula favour the control (SGL) curriculum and for the 'Surface' and 'Strategic' scales the confidence interval excludes zero.

Table 3.28: Mean ASSIST scale scores at T^2 and sensitivity analysis

Scale	Sample size Exp Start / Control effect	1- Exp. group (s.dev)	2- Cont. group (s.dev)	1-2	Effect size *d*	95% C.I	P
Surface Approach	33/20 33/31	16 (4.3)	14 (4.3)	2	-0.5	-0.1 to 1.0	0.06
Deep Approach	33/20 33/31	24 (3.5)	23 (2.9)	0.8	0.3	-0.3 to 0.9	0.16
Strategic Approach	32/20 31/31	22 (4.9)	23 (3.2)	-1.0	-0.3	-0.9 to 0.2	0.18
Surface Approach (M)	33/33 33/33	21 (6.1)	16 (5.5)	-4.7	-0.8	-0.3 to –1.4	0.002
Deep Approach (M)	32/32 33/33	21.9 (4.2)	22.4 (4.2)	-0.45	-0.1	-0.6 to 0.4	0.7
Strategic Approach (M)	33/33 33/33	18 (6.6)	21 (5.2)	-3.6	-0.7	-0.2 to –1.2	0.01

The second analysis of the ASSIST scales is to compare the difference in the change in mean scores for the two curricular between the first and second administration of the instrument. The null hypothesis being tested here is that Problem Based Learning has no effect on the development of 'better' approaches to learning.

Box 3.10: Formula used for calculating change effect size for the ASSIST instrument

$$\frac{(\text{Exp mean } T^1 - \text{Exp mean at } T^2) - (\text{Cont. mean at } T^1 - \text{Cont. mean at } T^2)}{(\text{S.Dev for control group at } T^1 + \text{S.Dev for control group at } T^2)/2}$$

Table 3.29 gives the results for the analysis of the comparison of changes in mean scores on the ASSIST scales between the first and second administration. The effect size calculations were carried out using the formula shown in box 3.10 above. On the 'Surface' scale the approach of students in both curricula improved (a decline in score indicating improvement on this scale) significantly. The improvement was greater in the experimental (PBL) curriculum a difference that was also significant. However, the effect size for the difference does not exclude zero. On the 'Deep' scale the approach of students in both curricula worsened significantly. The decline was less in the experimental (PBL) curriculum but the difference between the two curricula was

not statistically significant. The approach of student in both curricula worsened equally on the 'Strategic' scale although the difference was not statistically significant.

Table 3.29: Comparisons of changes in mean ASSIST scale scores T^1 -T^2

Scale	Group (n)	Mean at T1 (S.Dev)	Mean at T2 (S.Dev)	Difference in Change in mean	SGL avg. S.Dev	Effect size *d* (95% C.I)	P
Surface	PBL (20)	18.4 (4.4)	16.5 (4.3)	+0.4	4.6	-0.1 (-0.7 to 0.5)	0.7
	SGL (26)	16.0 (4.9)	14.5 (4.3)				
Deep	PBL (20)	25 (3.4)	24.4 (3.4)	-0.6	2.8	0.5 (-0.1 to 1.0)	0.5
	SGL (26)	25 (2.7)	23.5 (2.9)				
Strategic	PBL (20)	22.6 (5.3)	21.9 (5.1)	0	3.2	0 (-0.6 to 0.6)	0.3
	SGL (26)	23.7 (3.2)	23 (3.2)				

Follow-up questionnaire - The Self Directed Learning Readiness Scale (SDLRS)

Box 3.11: Scales and items in Self directed Learning Readiness Instrument

Subscale: Self management

- I manage my time well
- I am self-disciplined
- I am organised
- I set strict time frames
- I have good management skills
- I am methodical
- I am systematic in my learning
- I set specific times for my study
- I solve problems using a plan
- I prioritise my work
- I can be trusted to pursue my own learning
- I prefer to plan my own learning
- I am confident in my ability to search out information

Subscale desire for learning

- I want to learn new information
- I enjoy learning new information
- I have a need to learn
- I enjoy a challenge
- I enjoy studying
- I critically evaluate new ideas
- I like to gather facts before I make a decision
- I like to evaluate what I do
- I am open to new ideas
- I learn from my mistakes
- I need to know why
- When presented with a problem I cannot resolve I will ask for assistance

Subscale: Self - Control

- I prefer to set my own goals
- I like to make decisions for myself
- I am responsible for my own decisions/ actions
- I am in control of my life
- I have high personal standards
- I prefer to set my own learning goals
- I evaluate my own performance
- I am logical
- I am responsible
- I have high personal expectations
- I am able to focus on a problem
- I am aware of my own limitations
- I can find out information for my self
- I have high beliefs in my abilities
- I prefer to set my own criteria on which to evaluate my performance

The student follow-up questionnaire also contained an instrument designed to assess readiness for self -directed learning. Details of the development of the scale are given in part I of the report and items composing the scale and subscale are given in box 3.11 above. The internal consistency of the scales measured by Cronbach's Alpha were 0.86 for the 'Self Management' scale, 0.85 for the 'Desire for Learning' scale, 0.89 for the 'Self Control' scale, and 0.94 for the 'Total (SDLRS)' scales. Respondents are asked to rate their agreement with the statements on a scale from 1 'Strongly disagree' to 5 'Strongly Agree'. Scale totals are computed by adding the ratings given for each individual item.

Table 3.30 Comparison of means on the Self Directed Learning Readiness and subscales

Scale	Sample size Exp Start / Control effect	1- Exp. group (s.dev)	2- Cont. group (s.dev)	1-2	Effect size *d*	95% C.I	P
Self management	34/19 33/26	52 (7.5)	52 (5.2)	0.4	0.1	-0.5 to 0.6	0.4
Desire for Learning	34/19 33/25	52 (5.8)	52 (3.5)	0.3	0.1	-0.5 to 0.7	0.3
Self Control	34/17 33/25	64 (6.9)	63 (5.6)	0.6	0.1	-0.5 to 0.7	0.35
SDLRS Total	34/17 33/25	169 (18.8)	167 (12.1)	2.4	0.2	-0.4 to 0.8	0.3
Self management (M)	34/34 33/33	43 (11.5)	50 (5.9)	-6.6	-1.2	-1.7 to - 0.6	0.05
Desire for Learning (M)	34/34 33/33	47 (6.6)	50 (4.2)	-3	-0.7	-1.2 to -0.2	0.03
Self Control (M)	34/34 33/33	57 (9.1)	61 (7.0)	-3.7	-0.6	-1.1 to - 0.1	0.07
SDLRS Total (M)	34/34 33/33	147 (26.2)	161 (15.2)	-13.4	-0.9	-1.4 to - 0.4	0.01

Differences between the mean scores for each curriculum are given in table 3.30 above. The effect sizes on all the scales favour the experimental (PBL) curriculum but the confidence intervals do not exclude zero. The second set of scales are the results of the sensitivity analysis. The lowest score for each curriculum on each scale was used to impute the missing data. Analysis using this data produces effect sizes that favour the control (SGL) curriculum with confidence intervals that exclude zero.

The mean scores for the total scale on both curricular are above the threshold of 150 that is defined by the designers of the instrument as being indicative of being ready for self -directed learning (Fisher et al, 2001). However, the proportion of students above this threshold is different in the two curricula as the results in table 3.31 below indicate. The effect size favours the control (SGL) curriculum and the confidence interval only just crosses zero (0.004). However, the difference was not statistically significant.

Table 3.31: Student Follow-up questionnaire – Self Directed Learning Readiness Total scale score <150

Question	Sample size Exp Control Start /effect	1- Exp. group n (%)	2- Cont. group n (%)	Odds ratio	Effect size *d* (95% C.I)	P
SDLRS scale score <150	34/19 33/26	4 (21)	2 (8)	0.31	-0.6 (-1.2 to 0)	0.2

If as expected The 'Self Directed Learning Readiness Scale' and the ASSIST instrument are tapping into similar dispositions and practices a degree of correlation between individual students scores on the scales should be expected. Tests of the correlation between all of the scales on each instrument were carried out and the results are reported in table 3.32 below. The data for the ASSIST scores comes from the second post intervention administration of the instrument. The results indicate that the correlation's are in the expected direction. In four cases the correlation's are significant at the 0.05 level and in a further four cases significant at the 0.1 level. However on all scales with possible exception of the Strategic/SDLRS 'Self management' (39%) and 'Total' (27%) the proportion of variability explained by the association is quite low. The results do however provide additional confidence that the combination of the two instruments provides a valid indicator of students' practices and dispositions concerning their approach to learning.

Table 3.32: Correlation's between student scores on scales in the ASSIST and Self Directed Learning Readiness Instruments

SDLRS scales ASSIST scales	Self management	Desire for learning	Self control	Total
Surface	r = -0.29 p=0.06	r = -0.4 p=0.01	r= -0.111 p=0.2	r= -0.29 p=0.25
Deep	r=0.2 p=0.09	r= 0.25 p=0.04	r= 0.17 p=0.13	r= 0.24 =p=0.06
Strategic	r= 0.63 p=0.00	r=0.34 p=0.13	r=0.35 p=0.09	r= 0.52 p=0.00

Results teacher workload

There was a large amount of data missing from the first year of the teachers' self-collected data. The pattern of responses suggested that substantial parts of the data collection were completed retrospectively and according to a formula (e.g. 60 minutes per week tutorials), rather than based on the actual teaching activity carried out in a particular week. The data from year two appears to be of better quality. There are fewer missing weeks and the data patterns suggest that more attention was aid to accurate recording by the teachers. Only data from year two data was included in the analysis. The data for both years suggest that teachers were operating a digit preference (0,5) in their completion of the data forms e.g. time units reported as 30 minutes or 45 minutes.

The primary question for the analysis is to estimate the difference in the amount of time spent 'teaching' in the experimental (PBL) and control (SGL) curricula. Time spent by teachers from the control (SGL) curriculum providing tutorial support to students in the experimental (PBL) curriculum was included in the experimental (PBL) curriculum total. The results for each specific type of teaching and the total teaching time are given in table 3.33 below. There is not homogeneity of variance and the data are not normally distributed and the large standard deviations indicate that the mean is not a good measure of central tendency particularly in the control (SGL) curriculum. Data transformation did not result in normal distributions so analysis was carried out on raw data. For these reasons statistical analysis was carried out using both parametric and non-parametric tests and the P value given in table 3.33 below is derived from the non-parametric Mann-Whitney test.

The results refer to minutes per week per curriculum. The results appear to show that less teacher time is spent on preparation for, classroom contact with and tutorials for students in the experimental (PBL) curriculum. However, these data should be interpreted cautiously given the concerns about the quality of the data mentioned above and the non normal distribution. The

amount of time spent on preparation in the control (SGL) curriculum appears to be much larger than in the experimental (PBL) curriculum and this contributes a large part of the difference in total teaching time. There are a number of reasons why the high level of preparation seen in the control (SGL) curriculum, even if accurate, is probably atypical. One of the control (SGL) teachers was a new teacher and the another reported that she had revised much of the programme material extensively during the second year. In the control (SGL) curriculum data there were 13 weeks where the preparation time exceeded 1 whole working day (450 mins). Whilst feasible this is unlikely to be typical. If these weeks were removed from the control (SGL) curriculum total the difference in average preparation time was much reduced. Because of this 'atypicality' the preparation time and total teaching time will not be discussed further.

Table 3.33: Teachers workload mean number of minutes teaching per week

Scale	Sample size (weeks) Exp Start / Control effect	1- Exp . group (s.dev)	2- Cont. group (s.dev)	1-2	Effect size *d*	95% C.I	P
Preparation for teaching	54/42 57/48	46 (45.7)	270 (344.8)	-224	-0.6	-1.1 to -0.2	0.01
Classroom teaching	54/42 57/48	201 (183.3)	262 (143.0)	-61	-0.4	-0.8 to 0	0.00
Student tutorials	54/42 57/48	31 (36.7)	24 (34)	+7	0.2	-0.2 to 0.6	0.14
Total teaching time	54/42 57/48	278 (227.5)	556 (405.7)	278	-0.7	-1.1 to -0.3	0.00

The difference in the mean classroom teaching time of one hour per week may not appear of particular practical significance. However if considered over the length of the whole programme the difference in time appears to have more practical implications. The total classroom teaching time of the experimental (PBL) curriculum is approximately 24 working days (based on 7.5 hour working days) compared to approximately 33 days in the control (SGL) curriculum. Even given the questions about data quality this would seem to be quite a large difference that could have practical benefits for hard pressed teachers. However, this apparent difference can take on a different perspective if for example expressed in terms of number of hours classroom teaching

per student who completed the programme, 13hours 50 minutes in the control (SGL) curriculum compared to 15 hours 2 minutes in the experimental curriculum.

Discussion

Limitations of the study

This study was designed to provide answers to questions about the additional or relative impact of using a Problem Based Learning curriculum as an approach to teaching and learning in a continuing educational programme for nurses. The design and conduct of the study was driven by the aim of attaining the maximum internal validity possible within the context of a 'real-time' educational field trial. Internal validity is regarded as a prerequisite for external validity i.e. the generalisation of the findings to other settings and contexts (Cook & Campbell, 1979). Despite these intentions prevailing conditions in the field setting, the limitations of the researchers and of the resources available, may have resulted in planned or unplanned modifications in the design and implementation of the study. Of concern are modifications to the study design, the adequacy of the operationalization of the concepts into measurable outcomes and the methods of measurement used. Whilst the well designed and conducted randomised experiment is the optimal design for the minimisation of bias in studies designed to answer such questions (Boruch & Wortman, 1979; Torgerson & Torgerson, 2001), during the study planned and unplanned modifications to the study design may have undermined some of the assumptions inherent in the design.

Threats to internal validity - differences in baseline characteristics

Although the students were randomly allocated to each curricula, the baseline characteristics of the students in the two curricula were different in some respects (see table 2.3 in part II). However, it is not clear which, if any, of these characteristics are important prognostic factors for

the study outcomes used. A preliminary analysis of the association between these characteristics and the study outcomes estimated to what extent these differences may have influenced the results of the study. The difference between male and female student's rate of withdrawal from the programme has already been noted. Because the start number of males was comparatively small and became even smaller due to withdrawal, gender was omitted from the following analysis. As noted earlier, the difference between the two curricula in the proportions of students promoted after completing the programme is probably the result of the differences in starting grades. The characteristics and variables cross-tabulated are shown in table 3.34.

Table 3.34: Baseline characteristics and outcome variables tested for association

Characteristic	Outcome
Ever worked at another hospital	CEQ - All scales
	ASSIST - all scales
Undertaken previous post-registration education	Self Directed Learning Readiness - total
	Self Directed Learning Readiness - score <150(Φ)
	Student follow-up self rating of performance
Educated to degree level	Student follow-up rating of impact
	Student follow-up rating of impact <0 (Φ)
Attended school in the UK	Assignment mean scores
	Programme completion (Φ)
	Impact questions (participation in learning, teaching, practice development,- gained promotion, changed jobs) (Φ)
	Global Taught this way scale

The association between the characteristics and outcomes was assessed for the variables in table 3.34 using the statistical tests Phi (Φ) & ETA. Both are tests of the strength of association which produce a value between zero and one that is interpreted in the same way as Pearson's *r* i.e. zero =no association and one (+/-)= high level of association (Bryman & Cramer, 1995). Phi was used where both variables were nominal and ETA where one variable was nominal and the second interval (Kinnear & Gray, 2000). The results of the analysis should be interpreted with caution. Firstly because of the withdrawal rate the data in the outcome measures does not include all the cases in the starting sample. Secondly because of the comparatively small sample size even apparently strong levels of association may not be statistically significant (there were no cases where the level of association was more than 0.365). Thirdly given the number of

correlation's performed in the analysis (132) at the conventional level of statistical significance (1 in 20) approximately six statistically significant results will occur by chance alone. Three results were excluded because extreme cases (outliers) following the direction of association appeared to be distorting the results. The results given in table 3.35 are those for which the level of association was statistically significant at the 0.05 level or above.

Table 3.35: Statistically significant tests of association between characteristics of the sample and outcomes

Outcome	Worked at Hospital 4	Worked in another hospital
CEQ Assessment scale	ETA =0.324 P=0.02	
CEQ Independence scale	ETA =0.365 P=0.01	
CEQ Teaching Scale	ETA =0.321 P=0.02	
CEQ Goals & Standards scale	ETA =0.332 P=0.02	
Follow- up impact scale score <0	Φ = - 0.34 P=0.05	Φ = - 0.39 P=0.01
Follow-up impact scale score	ETA = 0.33 P=0.06	
Taken up a hobby		Φ = - 0.27 P=0.06
Been promoted since completing the course		Φ = - 0.27 P=0.06

For all the associations reported in table 3.35 the presence of the characteristic is negatively associated with the outcome (ETA does not report the direction of association in SPSS). That means, for example, working at hospital four is significantly associated with giving lower ratings on the course evaluation questionnaire (CEQ). The consistency of the associations between hospital four and the outcomes reported suggest that students working at this hospital give lower evaluations on these instruments. Thus if there were a greater proportion of these students in either curriculum this factor could be contributing to any differences in the mean outcome scores between the curriculum. However, although there was an imbalance in the proportion of students from hospital four in each curriculum at the start of the programme, because of withdrawals the proportion of students in each curriculum who worked at hospital four and completed the programme was the same (40%). Most of the data on the Course Evaluation Questionnaires and all of the data on the follow-up questionnaires was obtained only from students who completed the programme. Thus it would seem likely that any characteristic specifically associated with working in hospital four contributed equally to the mean scores of both curriculum groups.

The imbalance in the proportion of students in the two curricula who reported that they had worked in a hospital other than the one they in which they were currently employed persisted through to completion of the study although the difference in proportions between the two curricula narrowed to 10%. In all three cross tabulations of this variable reported in table 3.35 the direction of association was negative e.g. previously working in another hospital was associated with lower rate of taking up a new interest outside work. It seems reasonable to conclude that some unknown set of personal characteristics linked to the fact that people have worked elsewhere contributed to the difference between the two curricula on the outcomes 'taken up new interests' and 'rating of programme impact'. The higher proportion of males in the experimental (PBL) curriculum does seem likely to have made a contribution to the higher withdrawal rate seen. The literature on Problem Based Learning does not make reference to differences in response to Problem Based Learning related to gender. This appears worthy of further investigation.

Other threats to internal validity

A second potential threat to validity was the differential rate of withdrawal of students from the programme. It should be emphasised again that these were not students 'lost to follow-up' or from whom data was not collected but rather students who actually dropped out of their programme of study without completing it. As such this is an outcome measure in itself. However, the different withdrawal rates in the two curricula raises two issues for the data analysis. Firstly is it still reasonable to suppose that the outcome data were generated from a randomly allocated sample? An assumption that underlies most of the statistical tests used. There appears to be no solution to this problem other than to employ caution in the interpretation of the data. The second issue concerns the impact that differences in drop-out rates has on the average scores for other outcomes in a particular curriculum. The solution adopted was to undertake 'intention to treat' or sensitivity analysis for the majority of outcomes. What values should be used in such analysis is open to debate and challenge, but given the aim of

the study was to test whether Problem Based Learning added value a conservative approach was used. For the outcome categories 'approach to learning' and 'student satisfaction' it appears reasonable to assume that the students who withdrew were likely to be unsatisfied and/or to possess approaches to learning that were less congruent with Problem Based Learning. However, the use of conservative imputed missing values in the analysis of impact maybe more questionable.

A third threat to the internal validity of the study arises from a form of 'contamination between groups'. The field work setting meant that no additional limitations were applied to the admission of programme applicants other than the usual course requirements. A greater number of applicants than expected applied to undertake the programme who were working on the same ward in the same hospital. Because of this some individuals who worked on the same ward were randomly allocated to different curricula. The qualitative data from the observations and teachers accounts make it clear that this contributed to the anxiety of these particular students and their dissatisfaction with the experimental (PBL) curriculum. It seems reasonable to assume that if the there had not been a 'control' (SGL) curriculum running simultaneously the students in the experimental (PBL) curriculum may have been less anxious and dissatisfied.

Threats to construct validity - Confounding levels of construct - reliability of intervention

As noted in part II of the report there were some differences between teachers beliefs and teaching styles within the experimental (PBL) and control (SGL) curricula. Given the comparatively small sample size and rates of withdrawal it is difficult to explore the contribution of these differences to student outcomes. Some preliminary analysis comparing the mean scores for students of each teacher within each curriculum was undertaken on a number of variables. With regard to student satisfaction there were no statistically significant differences between

teachers in either curriculum on any of the scales in the Course Evaluation Questionnaires or in rates of student withdrawal. The difference in satisfaction scores between teachers in the control (SGL) curriculum on the follow-up questionnaire was statistically significant (F=7.96 P=0.003). On the impact indicators, mean assignment scores and follow-up questionnaire 'impact rating' there were no statistically significant differences between teachers in the control (SGL) curriculum. There were differences on these indicators between teachers in the experimental (PBL) curriculum. The effect size for the difference in the mean assignment score between the two teachers was d= 0.4 (95%C.I -0.3 to 1.5). On the students rating of impact of the programme on the follow-up questionnaire the effect size difference between the two teachers was d = 0.9 (95% C.I. 0.1 to 0.8, P=0.05). This limited analysis suggests that there may have been teacher effects interacting with curriculum effects on some outcome measures.

Threats to construct validity: Confounding levels of construct - Novelty effects

Inevitably a teaching and learning intervention develops in its implementation as teachers bring their own experience to the interpretation of the constructs in practice (i.e. Problem Based Learning'). It can be argued that this 'novelty' factor may diminish as teachers become more experienced in the practices required in the new programme. However, this argument should be cautiously applied to teaching and learning interventions as it appears to imply a narrow technical interpretation of teaching practice. In part II the consistent differences between the experimental (PBL) and control (SGL) curricula were explicated both in theory and practice. The fact that the teachers in the experimental (PBL) curriculum appeared to become less anxious and more relaxed as the programme progressed was noted as was the difference in practice between years one and two of the control (SGL) curriculum after a change of teacher. Given the comparatively small sample size and rates of withdrawal it is difficult to explore differences in student outcomes in the two years. Some exploratory analysis comparing the mean scores between curricula in each year was undertaken on a number of variables.

With regard to student satisfaction the data analysed presents a slightly contradictory picture. The difference in mean scores on the scales in the Course Evaluation Questionnaire was very similar and statistically significant in both years. However, the differences in the withdrawal rates from the two curricula declined from 43% to 28% between years one and two. Similarly the effect size for the difference in the mean scores on the follow-up questionnaire rating of satisfaction diminished from d= -2.6 in year one to d= -1.1 in year two. The difference between the two curricula in years one and two on the follow-up questionnaire impact scales were similar. The difference in the mean assignment score was different in year one and two. In year one the effect size was d=0.6 (95% C.I -0.3 to 1.5) i.e. in favour of the experimental (PBL) curriculum. In year two the effect size was -0.05 (95% CI -0.9 to 0.8) i.e. in favour of the control (SGL) curriculum. Based on this analysis it is suggested that any confounding levels of the construct that were present in either curriculum do not appear to have systematically affected the outcomes of the study.

Threats to construct validity - measurement issues

All of the categories of outcomes investigated i.e. impact on 'Skills Personal and Propositional Knowledge', 'Approaches to Learning', and 'Student Satisfaction' were operationalised and measured using a basket of indicators. Thus conclusions about impact in any one of these areas is based on a variety outcome measures and instruments decreasing the likelihood of problems associated with mono operation and method bias. The claim that Problem Based Learning aims to help students create better mental models of the world was probably only assessed to a limited extent by the programme assignments. In retrospect the activities carried out by the students for the video assessment were unlikely to provoke the explicit demonstration of capability in group or teamwork. The intra-item reliability of the instrument used to assess group work performance in the video sessions was not established independently of this study. With the exception of these two outcomes/instruments it is argued that the outcomes and instruments

used provide valid indicators of the individual dispositions and practices that comprise 'Skill, Personal and Propositional Knowledge', Approaches to Learning' and 'Satisfaction'.

Threats to statistical conclusion validity

The sample size in the study was small reducing the probability of obtaining results that were statistically significant and thus increasing the risk of type I error i.e. rejecting the null hypothesis when in fact it is true. Multiple dependent variables were analysed increasing the chance of a type II error i.e. not rejecting the null hypothesis when it is in fact false. However, several factors about the design and analysis of the study need to weighed against these problems. Firstly, the overall approach of the data analysis emphasises estimation rather than hypothesis testing. All of the data for all of the results including standard deviations, confidence intervals, reliability of scales, effect size and statistical significance are presented allowing the reader to make their own judgement on these issues. Secondly the sample size was sufficient to provide an acceptable probability of detecting effect sizes that, apriori, were thought necessary to demonstrate that Problem Based Learning made a substantial difference to student attainment. This was confirmed in the post-hoc sample size estimates. Thirdly the study was conceptualised as a phase II randomised trial in which a range of outcome measures and instruments are used in order to identify those which would be of most use in further definitive studies.

Issues in interpreting the results of the study

Researchers in education and other fields continue to debate the practical significance of an effect size. A general recommended guideline across disciplines is that d=0.2 (small effect), d=0.50 (moderate effect), and d=0.80 (large effect) (Cohen, 1988). It has been argued that an effect size of d=2.0 should be required where wholesale curriculum and organisational change is implied (Bloom, 1984). In the case of Problem Based Learning Colliver (2000) argued that d=1.0 should

be the minimum size of effect required to justify the considerable change that Problem Based Learning requires. Colliver's argument was accepted as the basis for the analysis of data in the study. However, it could also be argued that the important thing is how the effect size is derived. If the effect size is derived from a high quality randomised experiment then a difference of any size could be considered important. For example an effect size of d=0.2 would, all other things being equal, raise the pass rate on an exam from 50% to 58% (Coe, 2002b). A figure that many students and teachers would consider a worthwhile improvement. However, the quantification of extra physical properties e.g. capabilities, feelings or attitudes, cannot approach the level of measurement precision achieved in physical sciences (Nash, 2002). Thus measurement error, particularly of latent variables, such as those measured in this study, could account for many small to medium effect sizes found in a study (Gorard et al, 2002).

A second issue concerns the interpretation of confidence intervals. The main purpose of confidence intervals is to indicate the precision of the study sample estimates as population values. The values inside the confidence interval are simply more likely than those outside the confidence interval (Gardner & Altman, 1989). So if the stipulated effect size of interest is outside the confidence interval although it is not excluded it is unlikely. The 95% confidence interval can be interpreted as the range of values within which we can be 95% certain that the true population value will lie based on the sample estimate (Sackett et al, 2000). The confidence interval is affected by the sample size. In small studies such as this the width of the confidence intervals will be wide. Whatever the width of the confidence interval the population estimate is more likely to be nearer the centre of the interval rather than the extreme, with the sample estimate being the best indicator of the population value (Gardner & Altman, 1989). Therefore a different weighting can be given to studies depending on where the effect size d =0 i.e. no difference, falls in the 95% confidence interval.

Deriving conclusions

A key problem in social science is how to determine which conditional statements (or conclusions) are scientifically interesting and which are pure speculation based on little more than the private intuitions of the person making the assertions. Unfortunately there are no universally agreed firm criteria for making this judgement. The concern here is to identify the degree of usefulness of a particular conclusion or theory as an instrument for investigation and understanding of the phenomena in question rather than its relative degree of absolute truth. From this perspective the various findings from a study are neither absolutely true nor absolutely false but rather have differing levels of warrant (Gordon, 1993). The warrant for conclusions from any study is based on a combination of the underpinning theory, the experiential knowledge of those involved, the research design, the quality of the research and the interpretation of the results.

Given the limitations of the study the questions is what are warranted? In order to make explicit the basis for a given level of warrant for a particular conclusion the outcomes reported have been summarised along the various dimensions of validity considered in tables 3.37 to 3.39 (see pages 201-204). Regarding the interpretation of the confidence interval the approach adopted is that when the zero (i.e. no effect) in a 95% confidence interval is at the extreme end of a confidence interval this should be weighted more highly than when zero is near the centre of a confidence interval. If the sensitivity analysis confirms the direction of effect this is felt to give greater confidence in the direction of effect. The issue of validity of the outcome and instruments used is largely a question of judgement based on the use of the instruments in other studies. All of the outcome and instruments used have some validity otherwise they would not have been selected for use in the study. However, with the benefit of experience some outcomes and instruments appear more congruent with the questions addressed in the study than others. All of the instruments used were selected because they had reported satisfactory levels of internal reliability in previous studies. The reliability of the new follow-up questionnaires designed for this study was established on separate groups of students prior to their use on the participating

samples. However the internal and/or inter rater reliability of the scales when used is important in interpreting the reliability of the effects (Cook & Campbell, 1979).

A warrant level is derived from a consideration of the extent to which the result for each outcome is 'valid'. This is the figure given in the column headed 'warrant level' in each table. The criteria for the categories is given table 3.36 below. The categories represent the level of warrant for drawing conclusions to two broad questions. Categories one to three represent the level of warrant for the broad question 'Is there a real difference between the outcomes of students in the experimental (PBL) and control (SGL) curricula that is likely to be due to curriculum differences?' Categories four to six represent the level of warrant for a question which follows from the first namely 'if there is not an important difference in student outcomes between the two curricula does the result clearly indicate that an 'important' effect is unlikely?' By definition if a particular outcome receives a level of warrant for the first broad question a separate level of warrant for the second question is not required. Where an outcome has not been given a warrant level this indicates that there is insufficient justification to warrant any conclusion about effect.

Table 3.36: Criteria for levels of warrant

Question	Warrant	Definition
1. Effect size, no effect unlikely	1	Effect size, confidence interval, P-Value * or ** , Validity & reliability **, no effects of differences in baseline characteristics *
	2	Effect size, confidence interval, P-Value * or ** ,Validity & reliability * and/or effects of baseline characteristics
	3	Effect size confidence interval**, Validity & reliability* and/or effects of baseline characteristics, p-value does not exclude chance
2 Direction of effect uncertain	4	Confidence interval suggests important effects unlikely*, P-Value * or ** , Validity & reliability **, no effects of baseline characteristics
	5	Confidence interval suggests important effects unlikely *, P-Value * or ** ,Validity & reliability * and/or effects of baseline characteristics
	6	Confidence interval suggests important effects unlikely*, Validity & reliability * and/or effects of baseline characteristics and p-value does not exclude chance

Table 3.37: Summary, weighting of evidence and warrant for conclusions for category 'student satisfaction' and classroom teaching time

Outcome	Effect size - meets study definition of 'important' (+ or -) ** = *d*=1.0 + * = *d*= 0.4 to 0.9 (or equivalent)	Confidence interval suggests opposite effect unlikely **=yes excludes 0 * =0 at extreme of CI	Confidence interval suggest 'important effect *d*=1.0 Unlikely *=yes	Sensitivity analysis confirms direction of effect *= Yes	Probability of result by chance ** <0.01 * <0.05	Congruence of PBL outcome, study outcome, and instrument (Validity) ** High *moderate	Internal /inter-rater Reliability of instrument /measure in this study ** High *moderate	Imbalance in baseline characteristics affecting outcome * unlikely	Warrant level
Follow-up multi-item satisfaction	**	**	n/a	*	**	**	**		2
Global 'taught this way scale'	**	**	n/a	**	**	**	n/a		2
CEQ Scales								? contamination	
Workload	**	**	n/a	*	**	**	*	effect	2
Teaching	**	**	n/a	*	**	**	**		2
Goals	**	**	n/a	*	**	**	*		2
Independence			*	*		*	*		6
Assessment				*		*			
Withdrawal rate	*	**	n/a	n/a	**	**	n/a	? gender effect	2
classroom teaching time	*	*	n/a	not done	**	n/a	*		2

Table 3.38: Summary, weighting of evidence and warrant for conclusions category approaches to learning

Outcome	Effect size - meets study definition of 'important' (+ or -) ** = *d*=1.0 + * = *d*= 0.4 to 0.9 (or equivalent)	Confidence interval suggests opposite effect unlikely **=yes excludes 0 * =0 at extreme of CI	Confidence interval suggest 'important effect *d*=1.0 Unlikely *=yes	Sensitivity analysis confirms direction of effect *=yes	Probability of result by chance ** <0.01 * <0.05	Congruence of PBL outcome, study outcome, and instrument (Validity) ** High *moderate	Internal /inter-rater Reliability of instrument /measure in this study ** High *moderate	Imbalance in baseline characteristics affecting outcome * unlikely	Warrant level
Post intervention ASSIST Scales									
Surface	*	*		*		**	*		3
Deep			*			**	*		
Strategic		*	*	*		**	*		6
ASSIST scales - Change T1- T2									
Surface			*	n/a	*	**	*	*	4/5
Deep			*	n/a		**	*	*	6
Strategic				n/a		**	*	*	
Self Directed Learning Readiness Scales									
Self -management			*			**	**	*	6
Desire for learning			*			**	**	*	6
Self Control			*			**	**	*	6
SDLRS Total			*			**	**	*	6
SDLRS <150	*	*		*		**	**	*	3

Table 3.39: Summary and weighting of evidence, skills personal and propositional knowledge category

Outcome	Effect size - meets study definition of 'important' (+ or -) ** = *d*=1.0 + * = *d*= 0.4 to 0.9 (or equivalent)	Confidence interval suggests opposite effect unlikely **=yes excludes 0 * =0 at extreme of CI	Confidence interval suggest 'important effect *d*=1.0 Unlikely *=yes	Sensitivity analysis confirms direction of effect *=yes	Probability of result by chance ** <0.01 * <0.05	Congruence of PBL outcome, study outcome, and instrument (Validity) ** High *moderate	Internal /inter-rater Reliability of instrument /measure in this study ** High *moderate	Imbalance in baseline characteristics affecting outcome * unlikely	Warrant Level
Student rating of impact	*	*		*		*	**		3 (?distribution)
Student rating of impact <0	*	**		*		*	**		2/3 (p=0.09)
Video groupwork assessment				n/a			*		
Mean assignment scores						*	*	*	
Students rating of performance in practice	*	*		*		*	**	*	2/3 (p=0.07)
Manager rating of performance			*	*		*	**	*	6
Taken up hobby			*	n/a		*	n/a		6
Taken another course				n/a		*	n/a	*	
Participated in practice Devt.	*	*		n/a	*	*	n/a	*	2
Done teaching			*	n/a		*	n/a	*	6
Changed job			*	n/a		*	n/a	*	6
Gained promotion	*	*		n/a		*	n/a		3

Conclusions

Conclusions about the effects of Problem Based Learning in this study

None of the outcomes assessed were given the highest level of warrant. The various indicators used to measure 'student satisfaction' would have been given this level of warrant except for the concerns that were raised earlier about the effects of contamination between students in the two curricula. The results on the majority of the 'student satisfaction' indicators were given a level two warrant. The effect sizes on six out of the eight indicators used suggest 'important' effects in favour of the control (SGL) curriculum. Students in the experimental (PBL) curriculum were able to compare their experience, unfavourably in their view, with their colleagues in the control (SGL) curricula on a week by week basis. This is likely to have heightened their sense of dissatisfaction. Similarly the gender imbalance between the two curricula may also be having an effect on the withdrawal rate in particular. The evidence from the qualitative data combined with the fact that the effect sizes on many of the indicators were 'important' support the conclusion that a substantial portion of students did not like the form of Problem Based Learning offered. And for those that did not complete the programme, this was an important contributing factor to their decision to withdraw from the programme.

In terms of impact on skills, personal and propositional knowledge there is lower level of warrant (3) for concluding that students in the control (SGL) curriculum were more likely to perceive that the educational programme had changed their practice. There is a similar level of warrant for concluding that students in the experimental (PBL) curriculum rated their performance in practice higher than students in the control (SGL) curriculum. The level three warrant given to the finding that students in the control (SGL) curriculum appear more likely to be promoted reflects the fact that the effect size is important, the confidence interval excludes zero and the result was unlikely to have occurred by chance. However as noted earlier the groups

were not balanced at baseline in terms of their grade and this is likely to account for at least some of the difference found

There is a lower level of warrant (3) for concluding that students in the experimental (PBL) curriculum had higher scores on the 'surface' scale in the ASSIST instrument at the completion of the programme i.e. their approach to learning as measured on this scale was worse. However, the students in the experimental (PBL) curriculum also scored higher on this scale at the beginning of the programme, which is probably why the effect size for the change in scores on this scale between the first and second administration of the instrument favours the experimental (PBL) curriculum. There is also a level 3 warrant for concluding that a greater proportion of students in the experimental (PBL) curriculum did not meet the threshold score indicating readiness for self-directed learning after completing the programme. The ASSIST scores of the students who scored below 150 on the readiness for Self-Directed Learning Scale were however lower at the beginning of the study. Amongst this group the mean scores of the (4) students in the experimental (PBL) curriculum were lower than the (2) students in the control (SGL) curriculum. There is some suggestion therefore that the difference in the Self Directed Learning Readiness Scale score may also be due to baseline differences.

There is a level three warrant for concluding that time spent teaching in the classroom was lower in the experimental (PBL) curriculum. Using this indicator the 'cost' of Problem Based Learning curriculum could be argued to be lower than the control (SGL) curriculum. A cost benefit analysis might however indicate contradictory findings. It could be argued that the use of Problem Based Learning reduces cost (teacher workload) whilst appearing to have no detrimental affect on student attainment (as measured by assignment results -although this possibility of cannot be excluded by the results here). Alternatively it could be argued that the use of Problem Based Learning reduces cost (teacher workload) but results in a disbenefit in terms of student satisfaction and drop-out (see page 190).

There is a lower level six warrant for concluding that the results exclude 'important' effects on managers rating of performance, Readiness for Self -Directed Learning, changes in approaches to learning on the ASSIST instrument, changes outside work, taking on a teaching task in the workplace and changing jobs.

What conclusions from this study can be generalised?

Conclusions with acceptable levels of internal validity are highlighted above that is the results that underpin these conclusions have the necessary levels of internal validity. Internal validity maybe necessary for generalization but it is not sufficient. The important questions are therefore, to which population? and to what forms of Problem Based Learning can the conclusions be generalised? It could be argued that this was an *n* of one study i.e. one university or one form of Problem based Learning, or an *n* of five study i.e. with five teachers or five hospitals. Given the small sample size and particular setting caution is required in making generalisations to other groups, settings, and times. However this situation is not peculiar to this study but is a feature of research of all kinds. It is a particular strength of this study that the conclusions drawn have high levels of internal validity.

The 'thick' description of the sample, context and intervention that was given in part II of the report should enable the reader to draw their own conclusions about the generalisability of the study findings to their own particular circumstances. It is argued that the sample and context of this study is similar to that which prevails in continuing nursing education in England. It is argued that used in that context, the use of a Problem Based Learning curriculum similar to that used in this study is likely to produce similar effects. It is not clear whether the conclusions would be generalisable to pre-registration nursing education although the context in England is similar. If an 'authentic' Problem Based Learning curriculum was used in this context it would presumably

last the full three years of the pre-registration curriculum and this may have different effects. Student expectations and motivations may also be different at this level of education.

Comparison with other similar studies

Another way of assessing the generalisability of the conclusions of this study is to compare the results of this study with others in which the context, methods and/or outcomes measured are similar. The 'review of reviews' (Newman, 2003) carried out alongside this study as part of the Project on the Effectiveness of Problem Based Learning (PEPBL), includes several studies in which the context and /or outcomes are comparable to those found in this study[3]. Four studies of the effectiveness of Problem Based Learning in continuing medical education were included in the review (Chan et al, 1999; Doucet et al, 1998; Grol et al, 1989; Premi et al, 1994). Although some of these studies made reference to student satisfaction none of the measures used were sufficiently robust to be included in the review. In the study by Grol et al (1989) a Problem Based Learning curriculum acted as the control group. In the study the performance of students in the control group (PBL) did not improve as much as those in the intervention group. However, the outcome measure used was highly congruent with the intervention. In the case of the studies by Doucet et al (1998) and Premi et al (1994) the outcomes, which were multiple choice tests of knowledge, favoured the Problem Based Learning group and in the 'Premi' study the conference interval excluded zero. In both these studies and the study by Chan et al (1999) the change in outcomes between the first and second interventions also favoured the Problem Based Learning group. However in both the Doucet et al (1998)and Chan et al (1999) studies the intervention was about a very discrete topic (headache management and depression in the elderly respectively). This suggests that the models of Problem Based Learning used in these studies was different to that used in the study reported here. The Doucet et al (1998)and Premi et al (1994) studies were both controlled before and after designs, and thus selection bias cannot be excluded as a 'cause' of the results.

Only one study included in the 'review of reviews' measured student satisfaction using a method rigorous enough to be included in the review. This randomised controlled trial found that

[3] Readers are directed to the study report for more details of the methods and findings of this study . A copy of the study report can be found at http://www.hebes.mdx.ac.uk/teaching/Research/PEPBL/index.htm

students on the Problem Based Learning Curriculum rated their programme more highly on range of criteria including autonomy, involvement and innovation. The Problem Based Learning students rated their experience lower in terms of clarity and task orientation (Moore et al, 1994). However, in this study the Problem Based Learning programme lasted two or three years and the sample were Harvard Medical Students.

Two studies included in the 'review of reviews' assessed changes in student learning styles. The study by Moore and colleagues (Moore et al, 1994) tested students learning style preferences using the Preferred Learning Style Index and on both scales the change in effect sizes between the pre and post course questionnaire favoured Problem Based Learning. A second study included in the review compared student learning styles using an instrument called the Short Inventory of Approaches to Studying (Coles, 1985). Again the changes in effect sizes favour the Problem Based Learning group on all scales. The results in each study mirror those in this study. On the scales in which a high score indicates 'more effective' approaches to learning i.e. 'Discovery', 'Meaning', 'Versatility', 'Deep', 'Strategic', the mean scores were worse after the intervention in both the experimental and control groups but the decline in scores was less in the experimental Problem Based Learning curriculum. The consistency of these results suggests that this may be an area worth investigating further.

Only one study of Problem Based Learning in nursing education (Lewis & Tamblyn, 1987) was included in the 'review of reviews' discussed above. Additional effort was made outside of the 'reviews of reviews' to identify studies of the effectiveness of Problem Based Learning in Nursing and/or in continuing education. This effort cannot be considered as systematic and comprehensive as a systematic review. There appear to be very few studies of the effectiveness of Problem Based Learning in nursing education generally and even fewer in continuing nursing education. Only one study of the use of Problem Based Learning in continuing nursing education was identified (Peterson et al, 1999). There have been a number of evaluative studies of Problem Based Learning in undergraduate or pre-registration nursing programmes (Alavi,

1995; Andrews & Reece Jones, 1996; Lewis & Tamblyn, 1987; Rideout et al, 2002; White et al, 1999).

With the exception of the studies by Lewis and Tamblyn (1987) and Rideout and colleagues (2002) all of the these studies used single group pre & post test or post test only designs. Thus they did not address the question of the effectiveness of Problem Based Learning compared to any other educational intervention. The study by White and colleagues (1999) did in fact used a mixed design. The analysis of learning styles used a single group pre and post design whereas the results on student satisfaction were based on a non equivalent control group design. The study by Rideout and colleagues (2002) compared satisfaction and outcomes of nursing students from two different Universities in Canada, one of which used Problem Based Learning and thus is also a non equivalent control group design. In these two cases systematic differences in sample characteristics cannot be excluded as possible explanations for any differences found between the two groups. The study by Lewis & Tamblyn (1987) used a controlled before and after study design and measured the difference in students' performance at care planning. The outcomes favoured the Problem Based Learning students but the reliability of the outcome measures was not established and the assessors were not blind to the allocation status of the students. Despite these limitations in the design and methods of their studies the authors of all these studies draw favourable conclusions about the effects of Problem Based Learning even in some cases where their own results appear inconclusive.

Biley (1999) reported on a study that used participant observation and ethnographic interviews with a small sample (n=17) of nursing students. Students were in the final two years of a four-year course in the adult branch of a nursing programme that used a hybrid system of Problem Based Learning (a combination of Problem Based Learning and traditional educational methods such as skills laboratory work and lectures). The reaction of students' in his study appear to have been similar to the reactions of students in this study. He reported that students reacted negatively to the uncertainty caused by the lack of rigidly delineated knowledge and hard facts.

Students were also uncomfortable depending upon group collaboration to learn and were uncertain of what the facilitator's role should be. Additionally, students felt they were not learning anything of real importance and that the skills inherent in the Problem Based Learning process were not valuable. These factors led to students worrying about entering nursing with serious gaps in their knowledge.

A number of additional studies of the use of Problem Based Learning in continuing medical education were identified that were published after the completion of data collection in the review of Problem Based Learning in continuing medical education completed by Smits and colleagues (Smits et al, 2002c). Ozuah and colleagues (2001) studied the effects of Problem Based Learning in a paediatric course for qualified doctors. Using a control group interrupted time series design they measured the self directed work activity of doctors when they were having daily lectures or twice weekly Problem Based Learning sessions. During the Problem Based Learning session period students in this group spent greater time in self directed work. However, no outcome measures are reported so it is not clear what difference if any this made to student performance.

A series of studies of increasing rigour compared the effects of a Problem Based Learning programme with a 'mixed integrated lecture' programme in continuing education for occupational health physicians in the Netherlands. Using a single group pre and post -test design, the effects of using Problem Based Learning were compared on the topic of work rehabilitation guidelines for low back pain. The study participants had increased scores on knowledge and performance tests (Smits et al, 2002b). In the second study a controlled before and after design was used with a waiting list control group that received no intervention and the topic was the same. In this study the students in the Problem Based Learning group had higher scores on the post intervention knowledge (d=1.8 95%C.I 1.0 to 2.5) and performance tests (d=0.5 95% C.I -0.1 to 1.2) (Smits et al, 2002d). In the third study a randomised controlled trial design was used, the assessors were blind to the allocation status of the students and student satisfaction was also

assessed. In this study the topic was the management of mental health problems. The improvement in the knowledge and performance scores for the intervention and control groups were statistically significant (Smits et al, 2002a). However the effect size for the difference between the Problem Based Learning and control group was small on both assessments (d=0.1 and d=0.2 respectively) with confidence intervals that did not exclude zero. Moreover, the students in the Problem Based Learning curriculum were less satisfied than students in the control curriculum. The effect size for the difference in satisfaction immediately at the end of the study was d= -0.9 (95% C.I -1.37 to -0.4). In the follow-up survey the difference in the effect size was d= -1.0 (95% C.I -1.5 to -0.5).

On the basis of the review of studies outlines above it appears that the findings from this study are not at all dissimilar to those found in other studies using experimental or quasi-experimental designs, reliable and blinded assessment. This gives added confidence that the findings from this study maybe generalisable to other forms of Problem Based Learning and other subjects/disciplines. However there appear to be very few rigorous studies of learner satisfaction with which to draw comparison. Many texts on Problem Based Learning make reference to student anxiety, distress and even anger as anticipated reactions to the difficulties of the self-directed learning that is part and parcel of the Problem Based Learning approach (Crooks et al 2001). However, this is usually done in the context of preparing tutors and/or students for what to expect (see Woods 1995 for example), and seems to be regarded as phase that students will get through with appropriate support.

Why did the Problem Based Learning curriculum used in this study not lead to the predicted results ?

This study was undertaken both out of a desire to increase our understanding of the effectiveness of Problem Based Learning and make a contribution to the improvement of educational practice. A relevant model here is Poppers' evolutionary epistemology. In this model Popper

argues that we start from some problem, proceed to a tentative solution or theory which may be partly or wholly mistaken or in any case will be subject to error elimination, which may comprise of empirical testing. This creative process itself gives rise to new problems (Perkinson, 1982). The model could be applied to the development of Problem Based Learning and to this study of Problem Based Learning in several different ways. In the discussion that follows the model is applied in the following way. The initial problem is that we are unclear how best to improve the educational process such that students become more likely to develop and sustain the practices and dispositions associated with successful learning and high levels of performance. Problem Based Learning represents a trial solution and this study one of a large number of attempts at error elimination. Thus the message that should be taken from this study is not simply that Problem Based Learning does not work. Rather the problem is 'why in this study did Problem Based Learning not produce the kind of effects that were anticipated? A question for which there are a range of possible explanations.

One group of explanations concerns the context of the study. It could for example be argued that the characteristics of the sample were somehow exceptional. A number of studies have found that students with a low readiness for self-directed learning exhibited high levels of anxiety when exposed to a self- directed learning project (Grow, 1991; Wiley, 1983). Perhaps the students in this study had a low readiness for self-directed learning? If this is the case given that the sample were all qualified nurses it seems likely that a low level of readiness for self-directed learning will be found amongst qualified nurses generally. But the high levels of dissatisfaction reported here maybe specific to this sample and/or linked to the particular context of a part-time continuing education programme of comparatively limited duration. The fact that the students who appeared in the study appeared to be motivated by extrinsic factors at least as much as any intrinsic desire to learn does not appear particularly unique. Moreover, one of the rationales for Problem Based Learning is that it can harness extrinsic motivation in the pursuit of learning. These questions can only be addressed satisfactorily by the conduct of further high quality primary and secondary research.

A second group of explanations concerns the design and implementation of the Problem Based Learning curriculum used in the study. This was a particular form of Problem Based Learning which, it could be argued maybe deficient in some important aspect. For example, the Problem Based Learning tutors in the study may not have been as experienced and comfortable with Problem Based Learning as other teachers might be. It is certainly the case that the teachers themselves came to realise that the Problem Based Learning being practised in this study was beyond their previous experience and more challenging than they had expected. The Problem Based Learning tutors in the study were experienced teachers with a teaching qualification. If it is the case that the capability of the teachers is a key factor in Problem Based Learning then this has potentially far reaching implications for the generalisability of Problem Based Learning. It suggests that extensive preparation and training of even experienced qualified teachers sympathetic to the goals of Problem Based Learning is necessary before Problem Based Learning can be successfully implemented. The majority of Higher Education Teachers in the UK do not hold a teaching qualification and no doubt hold a wide range of views about teaching and learning.

It could also be argued that the preparation of the students for Problem Based Learning was insufficient. One question that arises here is how does one prepare for Problem Based Learning other than by doing Problem Based Learning? The very notion of 'preparing' for Problem Based Learning would appear to undermine the integrated conception of 'knowledge' that it is claimed underpins Problem Based Learning. Even if this logical disjunction was felt to be unproblematic, there remains the practical issue of how much time can realistically be devoted to preparing students for Problem Based Learning in a part time continuing education programme.

Whilst there are no doubt many different ways in which variant the Problem Based Learning practised in this study could be manipulated it is not sufficient simply to assert that this is or is not why the predicted affects were not seen. These assertions themselves require further empirical

study. One way in which this has been attempted in developing theories of Problem Based Learning is through causal modelling particularly by Schmidt and colleagues in work that has been developing for more than a decade (Gijselaers & Schmidt, 1990; Schmidt et al, 2003; Schmidt & Moust, 1995; Schmidt & Moust, 2000). Quite apart from any limitations of these particular studies it is highly contentious to claim that correlational data (as used in causal modelling) can be used to attribute causality (Wolfe, 1985). Factors identified by regression coefficients as important to the effectiveness of Problem Based Learning still require subsequent testing using experimental research designs.

A third set of possible explanations concerns the design and conduct of the research. It could be argued that the effects of Problem Based Learning are longer term in nature and thus studies in which the follow-up period is relatively short such as this one are unlikely to obtain evidence of effects. There are few longitudinal studies in Problem Based Learning. One study of the long term effects of Problem Based Learning compared graduates of the medical programme at McMaster University, which uses Problem Based Learning, with graduates of the medical programme at University of Toronto that uses a 'traditional curriculum'. The participants in the study had graduated between five and fifteen years before the follow-up survey (Shin et al, 1993). Participants were randomly selected but were it would appear a very specific subsection of the students who emerged from those institutions as only graduates who were at the time of the follow -up study working in General Practice in Ontario were included in the sampling frame. Moreover, the study design was a non-equivalent control group design, the study participants chose to go to either of the two institutions and thus the possibility of selection bias cannot be excluded. A long term follow-up of those medical students who took part in the randomised experiment comparing Problem Based Learning to the traditional curriculum at Harvard Medical School Programme was carried out ten approximately ten years after the students completed the programme (Peters et al, 2000).

The Canadian study found that graduates of McMaster scored more highly on some scales in a questionnaire designed to evaluate participants' management of hypertension and some of these differences were statistically significant (Shin et al, 1993). The Harvard study found that Problem Based Learning graduates scored more highly a group of scales that measured 'Humanism' but that there were no significant differences between the two groups on scales measuring lifelong learning attitudes and behaviours and satisfaction with their medical education (Peters et al, 2000). It is not easy to draw conclusions from the results of both studies as effect sizes are not given and neither is the data necessary to calculate them. The limitations of the design of the Canadian study mean that drawing the conclusion that differences between the two groups are due to Problem Based Learning is not supportable. The Harvard study suggests that there may be some effects of Problem Based Learning on attitudes that the authors call 'Humanism'. However, these effects were also detected in the initial evaluations of the programme (Block et al, 1994) and so appear to be immediate effects of Problem Based Learning that are sustained rather than effects that emerge some time after completion of the programme.

Another possible explanation in this category concerns that validity of the instruments used to measure the effects of Problem Based Learning. It seems reasonable to conclude that that the instruments used to measure 'Student Satisfaction' and 'Approaches to Learning' in this study had acceptable levels of validity. In relation to the programme design there are of course issues of the constructive alignment of these measures with the programme aims given that they were not part of the required programme assessments (Biggs, 2003)(Biggs 2003). The validity of the measures used to assess impact on 'Skills, Propositional and Personal Knowledge' is however open to question. This appears to be an issue in Higher Education in general rather than just in relation to Problem Based Learning (Cullen et al, 2002). Within Problem Based Learning it is generally concluded that the development of an evaluation system that is congruent with the purpose and philosophy of Problem Based Learning, provides valid reliable indicators of or proxies for developments in cognitive or performative practice, and that motivate learning is an ongoing development (Rideout, 2001).

Given the obvious need to keep the assessment and administrative burden for both teachers and students manageable it seems likely that such measures need to contain both formative and summative elements. It is surprising that Problem Based Learning programmes do not appear to make more use of the artefacts produced by students as a result of their Problem Based Learning activity for summative as well as formative evaluation. The development of subject or discipline specific instruments similar to the Progress Test used at University of Missouri - Kansas School of Medicine (Arnold & Willoughby, 1990), the University of Maastricht (Boshuizen et al, 1997), and the medical programme at McMaster University (Blake et al, 1996) maybe worth exploring further. Such tests appear to offer the promise of high levels of predictive power, student motivation and the tantalising prospect of allowing comparison of the effectiveness of different kinds of educational intervention. The use of traditional programme assignments or variants of this approach such as modified essay questions or the Triple Jump test are also worthy of further investigation.

What might be worth considering here is rather than focusing on the format of the assessment (the above are all varieties of assessment format) the focus should be on the content both in terms of the task that is set and the way the products are analysed. An interesting example of a different approach in the evaluation of Problem Based Learning is the work of Cindy Hmelo and colleagues. In a series of studies they assessed the development of cognitive skills in medical students by having the students develop hypothetico- inductive explanations to medical problems (Hmelo et al, 1997; Hmelo, 1998). An interesting example from outside the canon of Problem Based Learning literature is provided by Tynala (1999) who evaluated the effectiveness of a constructivist learning environment by analysing the development in students use of concepts in a series of different writing tasks using concept mapping. In an interesting parallel with this study the concepts were evaluated using the SOLO taxonomy. It is also interesting to note that both Hmelo and Tynala used quasi -experimental research designs adding weight to the argument that

such designs are both possible and necessary even where a more in depth, outcome specific method of assessment is used .

A problem with Problem Based Learning?

Returning to the evolutionary epistemology process let us suppose that in all the attempts at error elimination the same 'new' problem emerges. This is not necessarily such an outlandish supposition as it might at first seem. It has been demonstrated above that rigorous studies have consistently failed to detect 'important' effect sizes in favour of Problem Based Learning and/or where they have the form of Problem Based Learning practised diverges markedly from the Problem Based Learning model used in this study. This 'new' problem could be formulated thus. Why, despite various manipulations of the inner teaching and learning environment have the effects predicted for Problem Based Learning still not been demonstrated?

Pawson & Tilley (1997) argue that explanation comes from an understanding of mechanism acting in social contexts i.e. cause or regularity = mechanism + context. In the case of education interaction between student and teacher takes place in a particular classroom, in a particular university, in a particular subject discipline, in a particular professional context etc. They argue that mechanisms e.g. Problem Based Learning only lead to success if the context is conducive. Looked at from this perspective possible explanations lie outwith the practice of Problem Based Learning and/ or the methods used in its evaluation. In this study Problem Based Learning was conceptualised and described within a model of the inner teaching and learning environment that included Institutional and Disciplinary contexts. However this extension is novel in accounts of Problem Based Learning. Whilst it is the case that the implementation of Problem Based Learning in many large scale programmes makes reference to the importance of preparation not only of the teaching staff but also staff in clinical areas, in the archetypal model of Problem Based Learning the world outside the classroom makes its appearance only in the form of 'the entry

characteristics of students'. Little reference is made to organisational and institutional contexts other than those prevailing within the immediate educational environment. Government policy documents and/or those of regulatory and /or accrediting bodies are quoted as supporting the principles or aims of Problem Based Learning as if they were accurate representations of the reality of organizational and institutional cultures throughout a country, profession or subject/discipline.

One can find numerous examples of arguments that higher and professional education policy has and is in fact moving in a direction that is antithetical to the philosophies of learning that underpin Problem Based Learning (see Barnett, 1994 for example). In relation to nursing education Milligan (1999) points out that attempts to reclassify learning reflect power struggles that are far from resolved. Bechtel and colleagues (1999) point out that whilst Problem Based Learning may be congruent with ideals of learning in Universities, those in the world of nursing practice place emphasis on content driven competency based education. The description of the organisational and institutional contexts and their impact on teacher and students beliefs and expectations in this study illustrate this tension.

Problem Based Learning is often advocated because it appears to be supported by a number of theories or concepts of learning. However as Norman & Schmidt (1999) point out educational theory is still relatively speaking in its infancy. Furthermore, as Colliver (2002) points out most educational theory is heuristic rather than predictive in nature. From this perspective educationalist should be cautious when using such theory to design educational programmes. From a realist perspective it could be argued that the theoretical basis of Problem Based Learning is inadequate in that it lacks an adequately broad conception of the social structures necessary to achieve the goals that it sets out to achieve. Given the evidence that personal and organisational change and development are an interactive process (Boyatziz et al, 1995), this absence is likely to have practical consequences particularly where Problem Based Learning is used in professional education. For example students in this study spent three hours per week in a Problem Based

Learning tutorial compared with 30 hours per week in the workplace. Over the course of the programme that is roughly 90 hours in the inner teaching and learning environment and as opposed to 900 in the workplace environment. If the institutional context outside the inner teaching learning environment does not share the same philosophy of learning it seems highly unlikely that Problem Based Learning is on its own going to facilitate the development of the desired learning outcomes.

In the light of the analysis given here it seems reasonable to ask why is it that Problem Based Learning seems to be so popular amongst policy makers and to an increasing extent amongst educationalists. The most obvious points here is that any policy maker can use the term Problem Based Learning to mean almost anything. Similarly a teacher or curriculum developer can call any educational practice Problem Based Learning. As Savin-Baden (2000) indicates it is quite possible to implement a variant of Problem Based Learning that retains an essentially foundationalist view of knowledge and yet incorporates aspects of Problem Based Learning that apparently offer the promise of developing skills in group work, communication or presentation skills for example. In a variant of this argument Margetson (1998) argues that many medical school Problem Based Learning programmes have implemented only a limited version of Problem Based Learning which he calls 'the convenient peg' conception of Problem Based Learning in which the foundationalist model of knowledge remains intact.

The teachers in this study all in theory subscribed to the student centred approach to learning that was described in the programme handbook (not a document specific to the Problem Based Learning programme). However the analysis of their teaching beliefs showed that they in fact still held views about teaching and learning that are more commensurate with foundationalist teacher-centred approaches. In the control curriculum this was manifest in their teaching practice. However, the Problem Based Learning curriculum restricted the teachers options for adopting teacher centred practices. But what if it didn't? The difference between espoused theory and practice is presumably no less likely in the practice of Problem Based Learning than it is anywhere

else. The appeal to policy makers of Problem Based Learning is on one level straightforward. The claims that are made for Problem Based Learning include all the things that policy makers claim that they want from Higher Education including and perhaps especially 'skills'. At a second deeper level it could be argued that Problem Based Learning is at least acceptable to policy makers as it turns the spotlight on the teacher and focuses the responsibility for educational improvement squarely on the shoulders of practitioners.

Future developments in research on Problem Based Learning

None of the arguments made above should be interpreted to mean that efforts to develop greater understanding in the theory and practice of Problem Based Learning should not continue. They suggest that rather than prescribing Problem Based Learning as an effective approach for all contexts, and for all educational outcomes, if Problem Based Learning is to avoid being discarded as a passing fashion, there is a need to be more realistic about what can be achieved by the manipulation of the inner teaching and learning environment alone. Educational innovators need to extend their concept of the inner teaching and learning environment to include wider Institutional and Disciplinary contexts and devote as much energy to thinking about how they may contribute to changing them as they do changing what goes on in the classroom. Whilst such an argument in relation to Problem Based learning has been made before (see Savin-Baden, 2000 for example) it does not appear to have been enacted systematically in most treatments of the subject.

There are a number of coherent actions that appear to be important for the development of more adequate answers to the question of what forms of Problem Based Learning produce which outcomes for which students in which contexts? Firstly those involved in the development, research and implementation of Problem Based Learning need to make the forms of Problem Based Learning on offer more explicit. Again this argument has been made before

by Savin Baden (2000) and others. However, the contribution made here is to offer an example of how a particular conceptual framework (the inner teaching and learning and environment) can be used to facilitate this process and in doing so locate Problem Based Learning within a more general conception of teaching and learning in Higher Education. Secondly there is a need to develop and research better ways of evaluating student outcomes. In doing this it is argued that there is a need to move beyond discipline specific concerns and an apparent obsession with assessment formats. All forms of assessment represent approximations of complex extra physical phenomena. The issue is how adequate are they for the task in hand rather than their approximation to any idealised notion of 'truth'.

The two actions outlined above are important pre-requites for the third action. Despite the various different interpretations of Problem Based Learning it is probably one of the most coherent set of pedagogical practices currently on offer in Higher Education. This and the fact that it is in widespread use provide an excellent opportunity for large scale evaluation studies. For example there are a number of pre-registration nursing education programmes in England that use variants of Problem Based Learning and an even greater number which don't. All of these programmes share the same aims and objectives that are stipulated in frameworks set out by the Nursing and Midwifery Council and the Higher Education Funding Council For England. A natural experiment is therefore built into the educational system. Research funders, accreditation bodies, researchers and practitioners should work to develop ways of co-operating in the design and implementation of rigorous primary and secondary studies to take advantage of such naturally occurring opportunities, across disciplines, subjects and countries. The objective here is not to reinvent the wheel but to channel the wealth of innovative and creative talent in the field in such a way that the goal of improving students learning outcomes through a greater understanding of Problem Based Learning can be achieved.

References

Abrahamson S. 'Evaluation of continuing education in health professions: The state of the art', *Evaluation and the health professions,* 7, 1, (1984). pp 3-23

Alavi C. *Problem-based Learning in a health sciences curriculum,* (London, Routledge, 1995)

Albanese M. A. & Mitchell S. 'Problem-based Learning: A Review of Literature on its Outcomes and Implementation Issues', *Academic Medicine,* 68, 1, (1993). pp 52-81

Altman D. *Practical statistics for Medical research,* (London, Chapman Hall, 1991)

American Statistical Association. *How to conduct pretesting,* (Alexandria, ASA, 1997)

Andrews M. & Reece Jones P. 'Problem-based learning in an undergraduate nursing programme: a case study', *Journal of Advanced Nursing,* 23, 2, (1996). pp 357-365

Argyris C. & Schon D. *Theory in practice: Increasing Professional Effectiveness,* (San Francisco, Jossey Bass, 1974)

Arnold L. & Willoughby T. 'The quarterly profile test', *Academic Medicine,* 65, (1990). pp 515-516

Bailey J, Hoffman K., & Donaldson J. 'Dimensions of a PBL curriculum that enhance enculturation into a clinical community of practice', (Chicago, American Education Research Association, 2003).

Barnett R. *The limits of competence: knowledge, Higher education and Society,* (Buckingham, Society for Research in Higher Education/ Open University Press, 1994)

Barriball K., While A., & Norman I. 'Continuing professional education for qualified nurses: A review of the literature', *Journal of Advanced Nursing,* 17, (1992). pp 1129-1140

Barrows H. *Generic Problem-Based Learning Essentials, World Wide Web,* http://www.pbli.org (12 Mar. 2003)

Barrows H. S. 'A Taxonomy of Problem-based Learning Methods', *Medical Education,* 20, (1986). pp 481-486

Barrows H. S. 'Foreword', *Problem based learning: A research perspective on learning interactions,* Eds. D. Evenson & C. Hmelo, (Mahwah, Lawrence Erlbaum, 2000), pp vii-ix.

Barrows H. S. *Problem-Based Learning Applied to Medical Education,* (Springfield, Southern Illinois University School of Medicine., 2000)

Bechtel G. A., Davidhizar R., & Bradshaw M. J. 'Problem-based learning in a competency-based world', *Nurse Educ.Today,* 19, 3, (1999). pp 182-187

Behr A. 'Exploring the lecture method: An empirical study', *Studies in Higher Education,* 13, (1988). pp 189-199

Benner P. *From novice to expert: Excellence and power in clinical nursing,* (London, Addison-Wesley, 1984)

Benson G., Noesgaard C., & Drummond-Young M. 'Facilitating small group learning', *Transforming Nursing Education through Problem-Based Learning,* Ed. Rideout E, (Sudbury, Mass, Jones and Bartlett Publishers, 2001), pp 75-102.

Benson J. & Michael W. 'Basic principles of design and analysis in evaluation research: A twenty-year perspective on evaluation study design', *Encyclopaedia of education research,* Eds. H Mitzel & W Rabinowitz, (London, Collier MacMillan, 1982), pp 545-553.

Bereiter C. & Scardamalia M. 'Process and product in problem based learning', *Problem based learning: A research perspective on learning interactions,* Eds. D. Evenson & C. Hmelo, (Mahwah, Lawrence Erlbaum Associates, 2000), pp 185-195.

Berkson L. 'Problem-based learning: have the expectations been met?', *Academic Medicine,* 68, 10, (1993). pp S79-S88

Biggs J. *Teaching for quality at University,* (Buckingham, Society for Research in Higher Education/ Open University Press, 1999)

Biggs J. *Teaching for quality learning at university,* 2nd ed., (Buckingham, Open University press, 2003)

Biley F. 'Creating tension: undergraduate student nurses' responses to a problem-based learning curriculum', *Nurse Education Today,* 19, 7, (1999). pp 586-591

Black N., Murphy M., Lamping D., McKee M., Sanderson C., Askham J., & Marteau T. 'Consensus method development for clinical guidelines', *Health Services Research Methods: A guide to best practice,* Eds. N Black, J Brazier, R Fitzpatrick, & B Reeves, (London, BMJ Books, 1998), pp 199-211.

Black P. & Wiliam D. *Inside the black box: raising standards through classroom assessment,* (London, Kings College, 1998)

Blaikie N. *Designing Social Research,* (London, Polity Press, 2000)

Blake J., Norman GR, & Smith E. K. M. 'Report card from McMaster: Student evaluation at a problem-based medical school', *Lancet; :* 345, (1995). pp 899-902

Blake J., Norman G., Keane D., Mueller C., Cunnington J., & Didyk N. 'Introducing progress testing in McMaster University's problem based medical curriculum: Psychometric properties and effect on learning', *Academic Medicine,* 71, 9, (1996). pp 1002-1007

Bligh J. 'Problem based small group learning: an idea whose time has come', *Medical Education,* 311, (1995). pp 342-343

Block S. & Moore G. T. 'Project evaluation', *New Pathways to Medical Education,* Eds. D Tosteson, S Adelstein, & S Carver, (Cambridge MA, Harvard University Press, 1994), pp 114-122.

Block S. D., Style C. B., & Moore G. T. *an we teach humanism? A randomized controlled trial evaluating the acquisition of humanistic knowledge, attitudes and skills in the new pathway at Harvard medical school (Unpublished paper - obtained from author),* (1994).

Blomfield R. & Hardy S. 'Evidence-based Nursing Practice', *Evidence-based Practice: A critical appraisal,* Eds. L Trinder & S Reynolds, (Oxford, Blackwell, 2000), pp 111-137.

Bloom B. S. 'The 2 sigma problem: the search for methods of group instruction as effective as one to one tutoring', *Education Research,* 4, (1984). pp 4-16

Boruch R. F. & Wortman P. M. 'Implications of education evaluation for education policy', *Review of research in education,* Ed. Berliner D.C, (Washington DC, American Educational Research Association, 1979), pp 309-361.

Boruch R. F. *Randomized experiments for planning and evaluation,* (Thousand Oaks Ca., Sage

Publications, 1997)

Boshuizen H. P., Van Der Vleuten C., & Machiels-Bongaerts M. 'Measuring knowledge and clinical reasoning skills in a problem based curriculum', *Medical Education,* 31, (1997). pp 115-121

Boyatziz R., Cowen S., & Kolb D. 'Conclusion: What if learning were the purpose of education', *Innovation in Professional Education,* Eds. R Boyatzis, S Cowen, & D. A. Kolb, (San Francisco, Jossey Bass, 1995), pp 228-247.

Brookfield S. & Preskill S. *Discussion as a way of teaching: tools and techniques for democratic classrooms,* (San Francisco, Jossey Bass, 1995)

Brown C., Belfield C., & Field S. 'Cost effectiveness of continuing professional development in health care: a review of the evidence', *British Medical Journal,* 324, (2002). pp 652-655

Brown G., Bull J., & Pendlebury M. *Assessing student learning in higher education,* (London, Routledge, 1997)

Bruffee K. *Collaborative learning: Higher Education, interdependence and the authority of knowledge,* (Baltimore, The John Hopkins University Press, 1993)

Bruner J. *The process of education,* (Cambridge MA, Harvard University Press, 1977)

Bryman A. & Cramer D. *Quantitative data analysis for social scientists,* 2nd ed., (London, Routledge, 1995)

Burns I. & Glenn S. 'A new model for a new context?', *Problem-based learning in nursing,* Eds. S Glen & K Wilkie, (Basingstoke, Macmillan Press, 2000), pp 1-8.

Campbell M., Julious S., & Altman D. 'Estimating sample sizes for binary, ordered categorical and continuous outcomes in two group comparisons', *British Medical Journal;* 311, (1995). pp 1145-1148

Campbell M., Fitzpatrick R., Haines A., Kinmouth A., Sandercock P., Spiegelhalter D., & Tyrer P. 'Framework for the design and evaluation of complex interventions to improve health', *British Medical Journal,* 321, (2000). pp 694-696

Carper B. 'Fundamental patterns of knowing in Nursing', *Advances in Nursing Science,* 1, 1, (1978). pp 13-23

Case S. & Thomson L. 'Gender differences in student development: Examining life stories, career histories and learning plans', *Innovation in professional education,* Eds. R Boyatziz, S Cowen, & D Kolb, (San Francisco, Jossey Bass, 1995), pp 135-165.

Cervero R. M. 'Continuing professional education and behavioral change: A model for research and evaluation', *Journal of Continuing Nursing Education,* 16, 3, (1985). pp 85-88

Cervero R. M. *Effective continuing education for professionals,* (San Francisco, Jossey-Bass, 1988)

Chan D. H., Leclair K., & Kaczorowski J. 'Problem-based small-group learning via the Internet among community family physicians: a randomized controlled trial', *MD Comput.,* 16, 3, (1999). pp 54-58

Charlin B., Mann K., & Hansen P. 'The many faces of problem-based learning: a framework for understanding and comparison', *Medical Teacher,* 20, 4, (1998). pp 323-330

Clarke J. 'Evidence based nursing - an international perspective', (Southampton, University of Southampton, 1997).

Cleverley D. *Implementing Inquiry-Based Learning in Nursing,* (London, Routledge, 2003)

Coe R. 'What is an effect size ?', *Building Research Capacity,* 4, (2002). pp 6-8

Coe R. 'Its the effect size stupid: What effect size is and why it is important', 9-12-2002, (Exeter, Annual Conference of the British Educational Research Association, 2002).

Cohen E. 'Restructuring the classroom: Conditions for productive small groups', *Review of Educational Research,* 64, 1, (1994). pp 1-35

Cohen J. *Statistical Power Analysis for the behavioral sciences,* 2nd ed., (Hillsdale , NJ, Erlbaum, 1988)

Cohen L. & Manion L. *Research Methods in Education,* 3rd, (London, Routledge, 1989)

Coles C. R. 'Differences between conventional and problem-based curricula in their students approaches to studying', *Medical Education,* 19, (1985). pp 308-309

Colliver J. 'Effectiveness of Problem-based learning curricula: Research & theory', *Academic Medicine,* 75, 3, (2000). pp 259-266

Colliver J. 'Educational theory and medical education practice: A cautionary note for medical school faculty', *Academic Medicine,* 77, 12, (2002). pp 1217-1220

Colliver J. 'The research enterprise in medical education', *Teaching & Learning in Medicine,* 15, 3, (2003). pp 154-155

Cook T. D. & Campbell D. T. *Quasi-experimentation: design and analysis issues in field settings,* (Chicago, Rand McNally, 1979)

Cullen J., Hadjivassiliou E., Hamilton E., Kelleher J., Sommerlad E., & Stern E. *Review of current pedagogic research and practice in the fields of post-compulsory education and lifelong learning,* Tavistock Institute, 2002.

Davis D. A., Thomson O'Brien M. A., Freemantle N., Wolf F., Mazmanian P., & Taylor-Vaizey A. 'Impact of formal continuing medical education: do conferences, workshops, rounds and other traditional continuing education activities change physician behaviour or health care outcomes', *Journal of The American Medical Association,* 282, 9, (1999). pp 867-874

Davis W., Nairn R., Paine M., Anderson R., Oh M., & . 'Effects of expert and Non-expert facilitators on the small group process and on student performance', *Academic Medicine,* 67, 7, (1992). pp 470-474

Dawson S. 'Changes in the distance: Professional reappraise the meaning of management', *The Journal of General Management,* 20, (1994). pp 16-18

Delbecq A. & Van den Ven A. 'A group process model for problem identification and programme planning', *Journal of Applied Behavioral Sciences,* 7, (1971). pp 467-492

Department of Health. *Making a difference: strengthening the nursing, midwifery and health visiting contribution to Health and Healthcare,* HMSO, 1999.

Department of Health. *Working together - learning together: A framework for life long learning for the NHS,* Department of Health, 2001.

Desforges C. 'Familiar challenges and new approaches: necessary advances in theory and methods in research on teaching and learning', *The Desmond Nuttall/ Carfax Memorial Lecture,* 2000, (Southwell, British Educational Research Association, 2001).

Dewey J. *Logic, the theory of inquiry,* (New York, Holt, 1938)

Dolmans D. & Schmidt H. 'What directs self-directed learning in a problem based curriculum', *Problem Based Learning: A research perspective on learning Interactions,* Eds. D Evenson & C Hmelo , (Mahwah, New Jersey, Lawrence Earlbaum Associates, 2000), pp 251-262.

Dolmans D. 'The effectiveness of PBL. the debate continues: Some concerns about the BEME movement', *Med.Educ.,* 37, (2003). pp 1129-1130

Doucet M. D., Purdy R. A., Kaufman D. M., & Langille D. B. 'Comparison of problem-based learning and lecture format in continuing medical education on headache diagnosis and management', *Med.Educ.,* 32, 6, (1998). pp 590-596

Drummond M. F. *Principles of economic appraisal in health care,* (Oxford, Oxford University Press, 1980)

Du Florey C. 'Sample size for beginners', *British Medical Journal,* 306, (1993). pp 1181-1184

Egger M. & Smith G. 'Bias in location and selection of studies', *British Medical Journal,* 316, (1998). pp 61-66

Elstein A., Shulman L., & Sprafka S. *Medical problem solving: An analysis of clinical reasoning,* (Cambridge MA., Harvard University Press, 1978)

Engel C. E. 'Not Just a Method but a Way of Learning', *The challenge of problem based learning,* Eds. D Boud & G. P. Felletti, (London, Kogan Page, 1991), pp 22-33.

English National Board. *Creating Lifelong Learners: Partnerships for Care,* English National Board, 1994.

Entwistle N., Tait H., & McCune V. 'Patterns of response to an Approaches to Study Inventory across contrasting groups and contexts', *European Journal of the Psychology of Education,* 15, (2000). pp 33-34

Entwistle N. & McCune V. 'The conceptual bases of study strategy inventories in higher education', *Educational Psychology Review,* Forthcoming, (2003). pp

Eraut M. *Developing professional knowledge and competence,* (London, Falmer Press, 1994)

Facione N., Facione P., Sanches C., & Facione N. 'Critical Thinking Disposition as a Measure of Competent Clinical Judgment; The development of the California Critical Thinking Disposition Inventory', *Journal of Nursing Education,* 33, 8, (1994). pp 345-350

Facione P., Giancarlo C., Facione N., & Gainen J. 'The disposition toward critical thinking', *Journal of General Education,* 44, 1, (1995). pp

Facione P. A. *The California Critical Thinking Skills Test,* (Millbrae, California Academic Press, 1990)

Fallon M. & Pomfret A. 'Research on curriculum and instruction implementation', *Review of Educational Research,* 47, 1, (1977). pp 335-397

Farrow R. & Norman GR. 'The effectiveness of PBL. the debate continues: Is meta-analysis

helpful?', *Med.Educ.*, 37, (2003). pp 1131-1132

Feletti G. P. 'Reliability and validity studies on Modified Essay Questions', *Journal of Medical Education*, 55, (1980). pp 933-941

Feletti G. P. & Smith E. K. M. 'Modified essay questions: are they worth the effort?', *Medical Education*, 20, 126, (1986). pp 132-

Fisher A. & Scriven M. *Critical Thinking: Its definition and assessment*, (Point Reyes, Edgepress, 1997)

Fisher M., King J., & Tague G. 'Development of a self-directed learning readiness scale for nursing education', *Nurse Education Today*, 21, (2001). pp 516-525

Flanders N. A. *Analyzing teaching behavior*, (Reading Mass, Addison-Wesley, 1970)

Freire P. *Pedagogy of Freedom: Ethics, democracy and civil courage (P.Clarke, translation)*, (New York, Rowan & Littlefield, 1998)

Gardner M. & Altman D. 'Estimating with confidence', *Statistics with confidence - Confidence intervals and statistical guidelines*, Eds. M Gardner & D Altman, (London, BMJ Books, 1989), pp 3-6.

Garson G. D. *PA765 Statnotes: An online textbook, World Wide Web*, http://www2.chass.ncsu.edu/garson/pa765/statnote.htm (15 Nov. 2001)

Gijselaers W. & Schmidt H. 'Towards a causal model of student learning within the context of a problem based curriculum', *Innovation in medical education: An evaluation of its present status*, Eds. Z Nooman, H Schmidt, & E Ezzat, (New York, Springer, 1990), pp 95-114.

Gijselaers W. H. 'Connecting problem-based practices with educational theory', *Bringing problem-based learning to higher education: Theory and Practice*, Eds. L Wilkerson & W Gijselaers, (San Francisco, Jossey Bass, 1996), pp 13-22.

Goodstein L., Cooke P., & Goodstein J. 'The team orientation and behaviour inventory (TOBI):', *The 1983 annual for facilitators, trainers and consultants*, San Diego, University Associates, 1983), pp 108-114.

Gorard S., Prandy K., & Roberts K. *An introduction to the simple role of numbers in social science research*, ESRC Teaching & Learning Research Programme Research Capacity Building Network, 2002.

Gordon S. *The History and Philosophy of Social Science*, (London, Routledge, 1993)

Grol R., Mokkink H., Helsper-Lucas A., Tielens V., & Bulte J. 'Effects of the vocational training of general practice consultation skills and medical performance', *Med.Educ.*, 23, 6, (1989). pp 512-521

Grow G. 'Teaching learners to be self-directed', *Adult Education Quarterley*, 41, 3, (1991). pp 125-149

Habib F., Eshra D., Weaver J., Newcomer W., O'Donnell C., & Neff-Smith M. 'Problem based learning: a new approach for nursing education in Egypt', *Journal of Multicultural Nursing and Health*, 5, 3, (1999). pp 6-11

Hadwin A. 'Promoting self regulation: Examining the relationships between problem based learning in medicine and the strategic content learning approach', (New York, American Educational Research Association, 1996).

Hammersley M. 'Evidence-based practice in education and the contribution of educational

research', *Evidence -Based Practice: A critical appraisal,* Eds. L Trinder & S Reynolds, (Oxford, Blackwell Science, 2000), pp 163-183.

Harden R. & Davis M. 'The continuum of problem-based learning', *Medical Teacher,* 20, 4, (1998). pp 317-322

Harter S. 'A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components', *Developmental Psychology,* 17, (1981). pp 300-312

Hedges L. 'The basics of Meta-analysis', 2003, (Stockholm, 3rd Campbell Collaboration Colloquium, 2003).

Herbert R. & Bravo G. 'Development and validation of an evaluation instrument for medical students in Tutorials', *Academic Medicine.,* 71, 5, (1996). pp 488-494

Hmelo C., Gotterer G., & Bransford J.D. 'A theory-driven approach to assessing the cognitive effects of PBL', *Instructional Science,* 25, 6, (1997). pp 387-408

Hmelo C. 'Problem-based Learning: Effects on the early acquisition of Cognitive Skills in Medicine', *The Journal of the Learning Sciences,* 7, 2, (1998). pp 173-208

Hmelo C. & Evenson D. 'Introduction to Problem based learning: Gaining insights on learning interactions through multiple methods of enquiry', *Problem based learning A research perspective on learning interactions,* Eds. D. Evenson & C. Hmelo, (Mahwah, Lawrence Erlbaum, 2000), pp 1-18.

Hmelo C. 'The constructivist teacher', (Chicago, American Educational Research Association, 2003).

Howell Adams M., Whitlow J., Stover L, & Williams Johnson K. 'Critical Thinking as an educational outcome: An evaluation of current tools of measurement', *Nurse Educator,* 21, 3, (1996). pp 23-32

Hussey T. & Smith P. 'The trouble with learning outcomes', *Active Learning in Higher Education,* 3, 3, (2002). pp 220-233

Hutchinson L. 'Evaluating and researching the effectiveness of educational interventions', *British Medical Journal,* 318, (1999). pp 1267-1269

Jaques D. *Small Group teaching: A SEDA Induction pack,* SEDA, 1990.

Johnson R. *Marking essays: Guidance notes for tutors,* (Manchester, Clifton Press, 1993).

Johnson S. & Finucane P. 'The emergence of problem-based learning in medical education', *Journal of Evaluation in Clinical Practice,* 6, 3, (2000). pp 281-291

Kahney H. *Problem Solving: Current Issues,* 2nd, (Buckingham, Open University Press, 1993)

Kinnear P. & Gray C. *SPSS for windows made simple: release 10,* (Hove, Psychology Press, 2000)

Knox J. D. E. 'How to use modified essay questions', *Medical Teacher,* 2, 1, (1980). pp 20-24

Kolb D. A. *Experiential learning: Experience as the source of learning and development,* (Englewood Cliffs, Prentice Hall Inc, 1984)

Kutnic P., Jules V., & Layne A. *Gender and school achievement in the Caribbean,* Department for International Development, 1997.

Lave J. & Wenger E. *Situated learning, Legitimate peripheral participation,* (Cambridge, Cambridge University Press, 1991)

Lewis K. E. & Tamblyn R. M. 'The problem-based learning approach in baccalaureate nursing education: how effective is it?', *Nurs.Pap.,* 19, 2, (1987). pp 17-26

Long M. & Johnson T. *Influences on the Course Experience Questionnaire scales,* Australian Council for Educational Research, 1997.

Louis T., Lavori P., Bailar J., & Polansky M. 'Cross-over and self-controlled designs in clinical research', *New England Journal of Medicine,* 310, 1, (1984). pp 24-31

Margetson D. 'Education, Pedagogy and Problem Based Learning', *Research and Development in Higher Education,* Ed. A Viskovic, (Campbelltown, Higher Education Research and Development Society of Australia, 1993), pp 164-171.

Margetson D. 'What counts as Problem-based Learning', *Education for Health,* 11, 2, (1998). pp 193-201

Margetson D. 'The role of evaluation', *Problem-based Learning in a Health Sciences Curriculum,* Ed. Alavi C, (London, Routledge, 1995), pp 141-161.

Marks-Maran D. & Gail Thomas B. 'Assessment and evaluation in problem based learning', *Problem-based learning in Nursing A new model for a new context,* Eds. S Glen & K Wilkie, (Basingstoke, Macmillan Press, 2000), pp 127-150.

Maudsley G. 'Do we all mean the same thing by "problem-based learning"?: A review of the concepts and formulation of the ground rules', *Academic Medicine,* 74, 2, (1999). pp 178-185

McColl E., Jacoby A., Thomas L., Soutter J., Bamford C., & Steen N. 'Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients', *Health Technology Assessment,* 5, 31, (2001). pp

Mckeachie W., Pintrich P., Yi-Guand L., & Smith D. *Teaching & Learning in the college classroom: A review of the research literature,* National center for research to improve postsecondary teaching and learning, University of Michigan, 1986.

McKee M., Britton A., Black N., McPherson K., Sanderson C., & Bain C. 'Choosing between randomised and non randomised studies', *Health Services Research: A guide to best practice,* Eds. N Black, J Brazier, R Fitzpatrick, & B Reeves, (London, BMJ Books, 2000), pp 61-72.

Millies P. 'The relationship between a teachers's life and teaching', *Teacher Lore: Learning from our own experience,* Eds. W Schubert & W Ayres, (New York, Longman, 1992), pp 25-43.

Milligan F. 'Beyond the rhetoric of problem-based learning: emancipatory limits and links with andragogy', *Nurse Educ.Today,* 19, 7, (1999). pp 548-555

Moher D., Schulz K., & Altman D. 'The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials', *Lancet; :* 357, (2001). pp 1191-1194

Moore G. T., Block S. D., Style C. B., & Mitchell R. 'The influence of the New Pathway curriculum on Harvard medical students', *Acad.Med.,* 69, 12, (1994). pp 983-989

Mulhall A. 'Nursing research: our world not theirs ?', *Journal of Advanced Nursing,* 25, (1997). pp 969-976

Muller S. 'Physicians for the 21st century : Report of the Project Panel of the General Professional Education of the Physician and College Preparation for Medicine', *Journal of Medical Education,* 59, 2, (1984). pp

Myers Kelson A. & Distlehorst L. 'Groups in problem-based learning: Essential elements in theory and practice', *Problem-Based Learning: A Research Perspective on Learning Interactions,* Eds. D. Evenson & C. Hmelo, (Mahwah, Lawrence Erlbaum Associates, 2000), pp 167-184.

Nash R. 'A realist scheme for social explanation: on 'numbers and narratives'', *Building Research Capacity,* 4, (2002). pp 1-4

National Research Council. *Scientific research in education,* National Academy Press, 2002.

Nevo D. 'The conceptualization of educational evaluation: an analytical review of the literature.', *New Directions in Educational Evaluation,* Ed. E House, (Lewes, Falmer Press, 1986), pp 15-29.

Newman M., Papadopoulos I., & Sigsworth J. 'Barriers to evidence based practice', *Clinical Effectiveness in Nursing,* 2, (1998). pp 11-20

Newman M. *A pilot systematic review and Meta-analysis on the effectiveness of Problem Based Learning ,* Learning & Teaching Subject Network for Medicine, Dentistry and Veterinary Medicine, 2003.

NHS Centre for Reviews & Dissemination. *Putting Research into Practice,* 1999.

Nichols D. P. *My Coefficient (Alpha) is Negative!, World Wide Web,* http://www.ats.ucla.edu/stat/spss/library/negalpha.htm (13 Nov. 2002)

Norman G. 'Reliability and validity of some cognitive measures of clinical reasoning', *Teaching and Learning in Medicine,* 1, 4, (1989). pp 194-199

Norman G. 'What should be assessed', *The challenge of problem based learning,* Eds. D Boud & G. Feletti, (London, Kogan Page, 1991), pp 254-259.

Norman GR. 'RCT = results confounded and trivial: the perils of grand educational experiments', *Medical Education,* 37, (2003). pp 582-584

Norman G. & Schmidt H. 'The psychological basis of Problem-based learning: A review of the evidence', *Academic Medicine,* 67, 9, (1992). pp 557-564

Norman G. & Schmidt H. 'Of what practical use is a baby? Perspectives on educational research as a scientific enterprise', *Professional Education Research Questions,* 20, 3, (1999). pp 1-5

Norman GR & Schmidt H. 'Effectiveness of Problem-based learning curricula: theory, practice and paper darts', *Medical Education,* 34, (2000). pp 721-728

Nowlen P. *A new approach to continuing education for business and the professions: The performance model,* (New York, MacMillan, 1988)

Oakley A. *Experiments in knowing: gender and method in the social sciences,* (Cambridge, Polity Press, 2000)

Ozuah P., Curtis J., & Stein R. 'Impact of problem-based learning on residents' self directed learning', *Archives of Pediatrics and Adolescent Medicine,* 155, 6, (2001). pp 669-672

Painvin C, Neufeld V. R., & Norman G. ' The triple jump exercise- a structured measure of problem-solving and self-directed learning', *Proceedings of the 18th Conference of Research in Medical Education,* 1979, (Washington, 1979).

Pascarella E. & Terenzeni P. *How college affects students': Findings and insights from 20 years of research,* (San Francisco, Jossey-Bass, 1991)

Pask G. 'Learning strategies, teaching strategies and conceptual or learning style.', *Learning styles and strategies,* Ed. R Schmeck, (New York, Plenum Press, 1988), pp 83-100.

Patton B., Giffin K., & Patton E. *Decision making group interaction,* 3rd ed., (New York, Harper Row., 1989)

Pawson R. & Tilley N. *Realistic evaluation,* (London, Sage, 1997)

Perkinson H. 'Education and learning from our mistakes', *In pursuit of truth: Essays on the philosophy of Karl Popper on the occasion of his 80th Birthday,* Ed. P Levinson, (Atlantic Highlands NJ, Humanities Press Inc., 1982), pp 126-153.

Peters A., Greenberger-Rosovsky R., Crowder C., Block S., & Moore G. 'Long-term outcomes of the new pathway program at Harvard Medical School: A randomized controlled trial', *Academic Medicine,* 75, 5, (2000). pp 470-479

Peterson R., Hakendorf M., & Guscott T. 'Improving aged care education for Australian rural nurses using problem based learning', *Journal of Continuing Education in Nursing,* 30, 3, (1999). pp 120-127

Petrosino A., Boruch R., Rounding C., McDonald S., & Chalmers I. 'The Campbell Collaboration Social Psychological Educational; Criminological Trials Register (C2SPECTR) to facilitate the preparation and maintenance of systematic reviews of social and educational interventions', *Evaluation and research in education,* 14, 3&4, (2000). pp 206-217

Premi J., Shannon S., Hartwick K., Lamb S., Wakefield J., & Williams J. 'Practice-based small-group CME', *Acad.Med.,* 69, 10, (1994). pp 800-802

Prescott R., Gillespie W., Counsell C., Grant A., Ross S., Russell I., & Kiauka S. 'Factors that limit the number, quality and progress of randomised trials', *Health Services Research Methods: A guide to best practice,* Eds. N Black, J Brazier, R Fitzpatrick, & B Reeves, (London, BMJ Books, 1998), pp 86-97.

Quinn R., Faerman S., Thompson M., & McGrath M. *Becoming a master manager,* (New York, John Wiley and Sons, 1990)

Ramsden P. *Learning to teach in Higher Education,* (London, Routledge, 1992)

Redding N. 'Assessing the big outcomes', *Educational Leadership,* May, (1992). pp 49-53

Regehr G. & Norman G. 'Issues in cognitive psychology: implications for professional education', *Academic Medicine,* 71, (1996). pp 988-1001

Richardson J. 'A critical evaluation of the short form of the Approaches to Studying Inventory', *Psychology Teaching Review,* 1, (1992). pp 34-44

Rideout E. 'Evaluating student learning', *Transforming Nursing Education through Problem-based learning,* Ed. Rideout E, (Sudbury, Jones & Bartlett, 2001), pp 215-239.

Rideout E., England-Oxford V., Brown B., Fothergill-Bourbannais F., Ingram C., Benson G., Ross M., & Coates A. 'A comparison of problem-based and conventional curricula in nursing education', *Advances in Health Sciences Education,* 7, 1, (2002). pp 3-17

Rideout E. & Carpio B. 'The problem based learning model of nursing education', *Transforming Nursing Education through problem-based learning,* Ed. Rideout E, (Sudbury, Jones & Bartlett, 2001), pp 21-45.

Ritchie J. & Spencer L. 'Qualitative data analysis for applied policy research', *Analyzing qualitative data,* Eds. A. Bryman & R Burgess, (London, Routledge, 1994), pp 173-194.

Robson C. *Real World research: A resource for social scientists and practitioner researchers,* (Blackwell, Oxford, 1993)

Rolfe I. & Murphy L M. J. 'The interaction between clinical reasoning, group process and problem based learning assessment at the Newcastle Medical School', *Research & development in Problem Based Learning: Proceedings of the 2nd bi-annual conference of Australian Problem Based Learning Network,* Eds. M Ostwald & A Kinmouth, (Sydney, Charles Sturt University Press, 1994), pp 46-53.

Rudner L., Glass G., Evartt D., & Emery P. *Meta-Stat,* (Computer Software), (Maryland, LMP Associates, 2002).

Sackett D., Strauss S., Scott Richardson W., Rosenberg W., & Haynes R. *Evidence based medicine: How to practice and teach EBM,* 2nd, (Edinburgh, Churchill Livingstone, 2000)

Sampson E. & Marthas M. *Group process for the health professions,* (Albany NY, Delmar, 1990)

Saunders N., Engel C. E., & Feletti G. 'Clinical Supervisors Report', *Medical Teacher,* 4, 4, (1982). pp 151-154

Savin-Baden M. 'Group Dynamics and disjunction in problem-based contexts', *Problem-based learning in Nursing A new model for a new context,* Eds. S Glen & K Wilkie, (Basingstoke, MacMillan Press, 2000), pp 87-105.

Savin-Baden M. *Problem-based learning in Higher Education: Untold Stories,* (Buckingham, Society for Research in Higher Education and Open University Press, 2000)

Schmidt H. & Moust J. 'What makes a tutor effective ? A structural equations modelling approach to learning in problem- based curricula', *Academic Medicine,* 70, (1995). pp 708-714

Schmidt H. & Moust J. 'Factors affecting small-group tutorial learning: A review of research', *Problem Based Learning: A research perspective on learning interactions,* Eds. D Evenson & C Hmelo, (Mahwah, Lawrence Erlbaum Associates, 2000), pp 19-51.

Schmidt H, Bruysten R., & Soppe M. 'Generality of a causal model of Problem-based Learning: Further Evidence', (Chicago, American Educational Research Association, 2003).

Schmidt H. G. 'Problem -based learning: rationale and description', *Medical Education,* 17, (1983). pp 11-16

Schmidt H. G. 'Foundations of problem based learning; some explanatory notes', *Medical Education,* 27, (1993). pp 422-432

Schon D. *Educating the reflective practitioner,* (Oxford, Jossey Bass, 1987)

Shultz K., Chalmers I., Hayes R., & ltman D. 'Empirical evidence of bias: Dimensions of

methodological quality associated with estimates of treatment effects in controlled trials', *Journal of The American Medical Association,* 273, (1995). pp 408-412

Selltiz C., Wrightsman L., & Cook S. *Research methods in social relations,* (New York, Holt, Rinehart and Winston, 1976)

Senn S. *Cross over trials in clinical research,* (John Wiley, London, 1993)

Shin J., Haynes R., & Johnston M. 'Effect of problem based, self-directed undergraduate education on life long learning', *Canadian Medical Education Journal,* 148, (1993). pp 969-976

Silver M. & Wilkerson L. 'Effects of tutors with subject expertise on the problem-based tutorial process', *Academic Medicine,* 66, (1991). pp 298-300

Smits P., De Buisonje C., Verbeek J., van Dijke F., Metz J., & J ten Cate T. 'Is problem-based learning more effective than lecture-based learning in postgraduate medical education? A randomized trial', *Effectiveness of postgraduate education in occupational medicine,* Ed. P Smits, (Amsterdam, Netherlands School of Occupational Health, University of Amsterdam, 2002a), pp 79-96.

Smits P., van der weide W., & Verbeek J. 'Evaluation of a postgraduate educational programme for occupational physicians on work rehabilitation guidelines for patients with low backpain: A before and after comparison', *Effectiveness of postgraduate education in occupational medicine,* Ed. P Smits, (Amsterdam, Netherlands School of Occupational Health, University of Amsterdam, 2002b), pp 57-64.

Smits P., Verbeek J., & De Buisonje C. 'Problem based learning in continuing medical education: a review of controlled evaluation studies', *British Medical Journal,* 324, (2002c). pp 153-156

Smits P., Verbeek J., van Dijke F., Metz J., & J ten Cate T. 'Evaluation of a postgraduate educational programme for occupational physicians on work rehabilitation guidelines for patients with low back pain: A controlled study', *Effectiveness of postgraduate education in occupational medicine,* Ed. P Smits, (Amsterdam, Netherlands School of Occupational Health, University of Amsterdam, 2002d), pp 67-78.

Soppe M, Schmidt H., & Bruysten R. 'Influence of Problem Familiarity on Learning in a Problem-based Course', (Chicago, American Educational Research Association, 2003).

Suominen T., Kovasin M., & Ketola O. 'Nursing culture - some viewpoints', *Journal of Advanced Nursing,* 25, (1997). pp 186-190

Swanson D., Case S., & Van der Vleuten CPM. ' Strategies for student assessment', *The Challenge of Problem-based Learning,* Eds. D Boud & G. Feletti, (London, Kogan page, 1991), pp 260-274.

Tait H. & Entwistle N. 'Identifying students at risk through ineffective study strategies', *Higher Education,* 31, (1996). pp 97-116

Tate R. 'Experimental design', *Encyclopaedia of education research,* Eds. H Mitzel & W Rabinowitz, (London, Collier Macmillan, 1982), pp 553-561.

Taylor Fitzgibbon C. *Monitoring education: Indicators, quality and effectiveness,* (London, Cassell, 1996)

Thomas R. E. 'Problem-based learning: measurable outcomes', *Medical Education,* 31, (1997). pp 320-329

Thomson H., Hoskins R., Petticrew M., Ogilvie D., Craig N., Quinn T., & Lindsay G.

'Evaluating the health effects of social interventions', *British Medical Journal,* 328, (2004). pp 282-285

Torgerson C. & Torgerson D. 'The need for randomised controlled trials in educational research', *British Journal of Educational Studies,* 49, 3, (2001). pp 316-318

Torgerson C. 'Educational research and randomised trials', *Medical Education,* 36, (2002). pp 1002-1003

Tynjala P. 'Towards expert knowledge: A comparison between constructivist and a traditional learning environment in the University setting.', *International Journal of Educational Research,* 31, (1999). pp 357-442

Van den Bossche P, Gijbels D., & Dochy F. 'Does problem based learning educate problem solvers? A Meta-analysis on the effects of problem based learning', *7th EDINEB conference,* 2000, (Newport Beach USA, 2000).

Van Der Vleuten C. P. M. 'The assessment of professional competence: Developments, research and practical implications', *Advances in Health Sciences Education,* 1, (1996). pp 41-67

Verma M., Chatwal J., & Singh T. 'Reliability of essay type questions- effect of structuring', *Assessment in Education.,* 4, 2, (1997). pp 265-270

Vernon D. & Blake R. L. 'Does Problem-based Learning Work? A Meta-analysis of Evaluative Research', *Academic Medicine,* 69, 7, (1993). pp 550-563

Vygotsky L. *Mind in society: The development of higher psychological processes,* (Cambridge, MA, Harvard University Press, 1978)

Waddell D. L. 'The effects of continuing education on nursing practice: A Meta-analysis', *Journal of Continuing Nursing Education,* 22, 113, (1991). pp 118-

Walton H. J. & Matthews M. B. 'Essentials of Problem Based Learning', *Medical Education,* 23, (1989). pp 542-558

Walton J., Clark D., & Click N. 'An outcomes assessment of a hybrid-PBL course in treatment planning', *Journal of Dental Education,* (1997). pp 361-367

Watkins K. & Marsick V. *Sculpting the learning organizations: Lessons in the art and science of systematic change,* (San Francisco, Jossey Bass, 1993)

Wenger E. *Communities of practice: Learning, Meaning, and Identity,* (Cambridge, Cambridge University Press, 1998)

White M., Amos E., & Khouzekanami K. 'Problem-Based learning: An outcomes study', *Nurse Educator,* 24, 2, (1999). pp 33-36

Wiley K. 'Effects of a self-directed learning project and preference for structure on self-directed learning readiness', *Nursing Research,* 32, 3, (1983). pp 181-185

Wilkes M. & Bligh J. 'Evaluating educational interventions', *British Medical Journal,* 318, (1999). pp 1260-1272

Wilkie K. 'The nature of Problem-based learning', *Problem based learning in nursing,* Eds. S Glen & K Wilkie, (Basingstoke, Macmillan Press, 2000), pp 11-34.

Wolf F. 'Problem-based learning and Meta-analysis: can we see the forest through the trees?', *Academic Medicine,* 68, 7, (1993). pp 542-545

Wolfe L. 'Application of causal models in higher education', *Higher Education: handbook of theory and research,* Ed. J Smart, (New York, Agathon, 1985), pp 87-133.

Wolff A. *Problem Based Learning. The role of the tutor: A resource guide for faculty,* (Vancouver, Wollf Consulting Ltd, 2000)

Wolff A. & Rideout E. 'The Faculty role in Problem Based Learning', *Transforming Nursing Education Through Problem-Based Learning,* Ed. Rideout E, (Sudbury MA, Jones and Bartlett, 2001), pp 193-213.

Wood I. 'The effects of continuing professional education on the clinical practice of nurses: A review of the literature', *International Journal of Nursing Studies,* 35, (1998). pp 125-131

Woods D. R. *Problem-based learning: helping your students gain the most from PBL,* (Hamilton CA: DR Woods., 1995)

Woodward C. & Ferrier B. 'Perspectives of graduates two or five years after graduation from a three year medical school', *Journal of Medical Education,* 57, (1982). pp 294-302

Woodward C. 'What can we learn from programme evaluation studies in medical education', *The challenge of problem based learning,* Eds. D Boud & G. Feletti, (London, Kogan page, 1997), pp 294-307.

World Bank. *World Development report 1993: Investing in Health,* Oxford University Press, 1993.

World Health Organization. *Increasing the relevance of Education for Health professionals: Report of a WHO study group on Problem Solving Education for Health professionals,* WHO, 1993.

Wragg E. C. *An introduction to classroom observation 2nd edition,* (London, Routledge., 1999)

Zhang X. & Zhang B. 'To apply problem-based learning teaching model install course of 'basic skills and tactics for problem solving'', *Shanxi Journal of Nursing,* 14, 1, (2000). pp 25-26

Zimmerman B. 'Models of self regulated learning and academic achievement', *Self-regulated learning and academic achievement: Theory, Research and Practice,* Eds. B Zimmerman & D Schunk, (New York, Springer-Verglag, 1989), pp 1-25.