This thesis is submitted in fulfillment of the requirements for the degree award of Doctor of Philosophy (PhD)

PhD Title: Discovering Effective Pedagogical and Evaluation Approaches for Learning Objects in Medical Education

SID NUMBER: 4031881

NAME: DAVINA SANDRA CALBRAITH

AFFILIATION: UNIVERSITY OF NOTTINGHAM

DATE OF SUBMISSION: FEBRUARY 2011

ABSTRACT

In 2004 unexplained pedagogical barriers were limiting Learning Object (LO) development. Few reference points existed preventing the formation of specific pedagogical questions as to *the nature of* these barriers - hence this PhD's rationale. This thesis 'uncovers' the most effective pedagogical and evaluation/assessment* approaches for LO design in Medical Education, and the underlying principles *within* these approaches – i.e. *what* is effective, and *why*.

To determine *why* certain approaches are effective observation/interview/usability studies were performed using grounded theory to generate hypotheses (1A Participants n=57). To verify 1A findings, this process was replicated using different sites/samples in Phase 2 (Eastern/Midlands, n=72). To determine *what* was most effective, systematic reviews using a purpose-built design were undertaken with additional questions on pedagogy and evaluation/assessment* components (1B Studies n=222). Approaches identified as 'effective' according to statistics, SCIE and my own rigor scoring systems were tested blind in two locations (Eastern/Midlands) with different samples under a null hypothesis (i.e. 'Each approach will score no differently to any other', Phase 2 participants n=72). This was further developed by replicating this process via *mobile* delivery.

Section 1A generated over a hundred hypotheses. In Section 1B, two existing approaches scored consistently high. Phase 2 produced the same hypotheses/approaches when submitted to the blinded observation/interview/usability process thus tight theme linkage resulted in rigorous theory and empirical data. The two top-performing 1B approaches scored high resulting in the possible existence of generic principles. When replicating 1A, 1B and Phase 2 for mobile delivery, the existence of generic principles was verified and a possible model for practice formed.

In summary, this thesis underlines the importance of learner input and how learners' perceptions form an essential part of the LO learning process. It discovers original generic principles for both desktop and mobile formats, highlights how branch and loop learning systems are necessary for learner customisation, and provides new knowledge verifying Wiley's molecular LO analogy.

* In this thesis many types of evaluation approaches are tested. These are called 'evaluation approaches' by the authors that created them. However, in some disciplines the term 'evaluation' is viewed as being interchangeable with the term 'assessment'. For this reason explanatory footnotes will be given throughout where necessary.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my husband, Steve Calbraith, who allowed me the necessary 'space' to undertake this work and without whose fantastic support this PhD would not have been possible. I would like to thank the Eastern R & D department, Dr Iain Stemp, and Dr Neil Harris who answered my ethical and statistical questions. I also would like to thank Trina Medlicott, whose administrative knowledge made this process a little easier and Prof Woody Caan, Ian, Carole, Martin and Julie who encouraged me during the write-up period. Lastly (but not least) I would like to thank all the willing participants for the various studies of which there are too numerous to mention individually here. I was particularly encouraged by the genuine interest and excitement regarding this work.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1. INTRODUCTION

1. INTRODUCTION

A good introduction to any thesis must surely include both how the idea was formed and why it was undertaken. This chapter will therefore outline my thinking in these respects. A definition of both 'Learning objects' (LOs) and 'Learning object learning' will be given, together with thoughts on how and why this research should be performed, the broad aim of the thesis, and how it will be set out.

1.1 The Idea for This Thesis

My interest in this area was originally kindled during the exponential rise of Learning Objects (LOs) during the 1990s when the concept first emerged. The stirrings of an idea started to materialize concerning how the development of LO pedagogy and evaluation appeared inextricably linked. An initial aim was to find evaluation approaches (together with their underlying pedagogy) for learning objects that worked in *practice*. At this stage experts in the field were uncertain regarding how these areas should be approached and even where a starting point should be. This further consolidated my notion that the topic may have some doctoral mileage.

The increasing interest in e-learning and LOs led to many new initiatives, government funding and hot debates between pedagogical and technical stances. Traditionally Information Technology (IT) has taken an Instructional Design (ISD) or Constructivist-type approach to research. Conversely, education has usually taken a more formative or narrative approach with regard to the reliability and validity of research undertaken. Since both learning objects (LOs) and E-learning need to be firmly rooted in both education and IT (and neither approach on their own appears adequate) it became evident that advancement of adult learning theory (or more appropriately pedagogy) within e-learning may require a new approach. The relative immaturity of academic rigor within e-learning compelled many educators and researchers to take a fresh look at LO and e-learning approaches - resulting in the gradual realisation that a more systematic approach was required. This apparent link between the need for systematic approaches and pedagogy further strengthened my interest.

At this time, many thought that LO adoption would have a huge impact on both e-learning and education (Merkow, 2002). Some had hoped that LOs would form an 'essential e-learning infrastructure' due to the fact that they had initially had 'significant impact in white papers' (Friesen 2006). In practice, the immediacy and flexibility of on-line learning objects (LOs) was seen both in healthcare and education (according to my experience) to be a great benefit for quick, timely and cost-effective learning. Unfortunately, many educators did not know where to start or what foundation to base their research on and largely relied on 'well.... it seems to work' type assumptions. Furthermore, despite the surge in popularity of *reusable* LOs (RLOs) and the emergence of digital repositories, most educators remained unsure regarding where to locate LOs and usually ended up devising their own.

In 2004, my initial interest was *not* as may have been expected at this point concerned with the *reusability* of LOs, but instead it was concerned the issues and processes that appeared to be preventing further development of *pedagogy and evaluation within LOs*. Of particular interest were the underlying reasons for i) why some LOs appeared to work in practice very effectively; ii) why others didn't; and iii) the complete lack of any kind of baseline on which to adequately assess them.

Furthermore, I had developed an intuitive belief that learning objects may contain *intrinsic potential* concerning their educational worth and flexible capability. I reasoned that systematic charting of research, evaluation and pedagogical evidence (together with reasoned theory) was theoretically possible, and this in turn could produce a more robust evidence base on which to build. Having previously mentioned that experts were unclear on where to start when studying LOs, work of this kind would give a defined and reasoned 'starting point', and at the very least provide further knowledge concerning LO research, evaluation and pedagogy. Thus, in 2004, the topic for this thesis was conceived.

1.2 Learning Object Definition

It is important to elucidate the central premise on which my thesis is based on, thus the definition of the Learning Objects (LOs) used in this thesis will be made clear. On looking briefly at the literature it is evident that many LO definitions contain similar ingredients, yet there is *not one definition* that is *universally agreed upon*.

Barron (2000) defines LOs as 'a new model from digital learning' where learning is content-free, is capable of being used in different systems and can be reused and continuously updated. Wiley's (2002a) definition is larger including '*any* digital resource that can be reused to support learning'. McGreal (2004, p13) expands this further: "any reusable digital resource that is encapsulated in *a lesson or assemblage of lessons grouped in units, modules, courses, and even programmes*". IEEE's (2002, p6) definition is even 'larger': "*Any entity,* digital or non-digital, that may be used for learning, *education or training*". Conversely, Leeder et al (2002) define LOs as 'Small chunks of interactive e-learning', whilst Howard-Rose and Harrigan (2003, p1). define them specifically as: "interactive computer programs... designed for students to use in a 15 minute to three hour time span as a mechanism to help them learn". Conversely, Darby (2003, p2) describes them as "the smallest element within an online course that defines a learning activity". So which of these two stances are correct?

No consensus is forthcoming and with such fundamental differences concerning the optimal size and nature of LOs, the first step must surely now be to arrive at a robust and reliable definition. Koper (2001) took up this challenge and tried to make a distinction between 'context resources' and the 'learning content design' (This distinction between content and context perhaps paved the way for Generative LOs to later follow). Koper concluded that "the lack of a precise and agreed upon definition of learning objects, besides making any serious study seem fuzzy and ill-planned, also limits productive dialogue and theoretical understanding of the application of learning objects in real-world implementations" (p45). Thus Koper noted the problem but did not fully address the challenge. As a result, Koper leaves the LO definition deliberately large: "A fundamental idea is that a learning object can stand on its own and may be reused" (Koper 2001, p45). This does not help towards a fuller more focused LO definition thus it is clear that the concepts in and around LOs, their development and utilisation are complicated, largely unknown and untested. As the jury is still 'out', I will ascribe to the larger definition of LOs until greater insight is gained.

At first glance, the IEEE's definition would be the most applicable for my thesis - its wide-ranging scope fits with the thesis's ideal of collecting *all* existing empirical evidence. However, this would require

unlimited time and resources and is therefore beyond the scope of this thesis. Since future LO development is sought (and non-digital entities were unlikely to tell me specifically how digital ones operate or how they can be used/developed) non-digital entities will be excluded. So little is presently known regarding LOs in terms of the reliability and validity of specific processes and practices. Hence, the definition has to be *large enough* to gather empirical data yet *'focused' enough* to make the learning explicit - i.e. not indirectly 'water down' project results by default. Although there is no consensus concerning size or nature, there appears to be an underlying assumption here that each topic is clearly defined. For this reason, Weller's et al's (2003, p2) definition was considered: 'A learning object addresses one clearly identifiable topic or learning outcome and has the potential to be reused in different contexts'. However, this does not specifically indicate digital resources but Howard-Rose and Harrigan's (CLOE, 2003 p1) definition does: "A learning object is any digital entity designed to meet a specific learning outcome that can be reused to support learning".

Since each of these definitions contained parts of what I believe LOs truly are, it was decided that an amalgamation of IEEE, Howard-Rose and Harrigan's, and Weller et al's definition would be used together with my own additions. The working definition of LOs for the purposes of this thesis is therefore: "any *digital* resource (reusable or otherwise) that has a clear learning objective or identifiable topic that may be used for multiple or different uses within learning, education or training in any course/curriculum".

1.3 A Definition of LO Learning

Robertson & Fluck (2004, p1) believe that 'good' learning objects are "fast moving, appear real, include colourful graphics, and use minimal amounts of text". Although this hints at what the learning may incorporate, the concept and definition of 'LO learning' does not currently exist. Much of the current research and literature seems unaware of the possibility that LOs may have intrinsic educational worth besides their reusability, and as such the value of finding how LO learning is best delivered and evaluated may be underestimated.

As the concept of 'learning object learning' is original, and no definition exists, for the purpose of my thesis the definition will be taken to be 'any learning involving any *digital* resource (reusable or otherwise) that has a clear learning objective or identifiable topic that may be used for multiple or different uses within learning, education or training in any course/curriculum'. Although parts of this thesis may necessitate taking a wider look at other disciplines, my thesis will place emphasis on and be performed in the areas of Medicine and Nursing. It will encompass a multi-site approach and results will be compared. Once complete, this thesis will revisit the phrase 'LO learning' in order to provide a fuller and more reasoned definition.

An Evidence Base for LO Learning

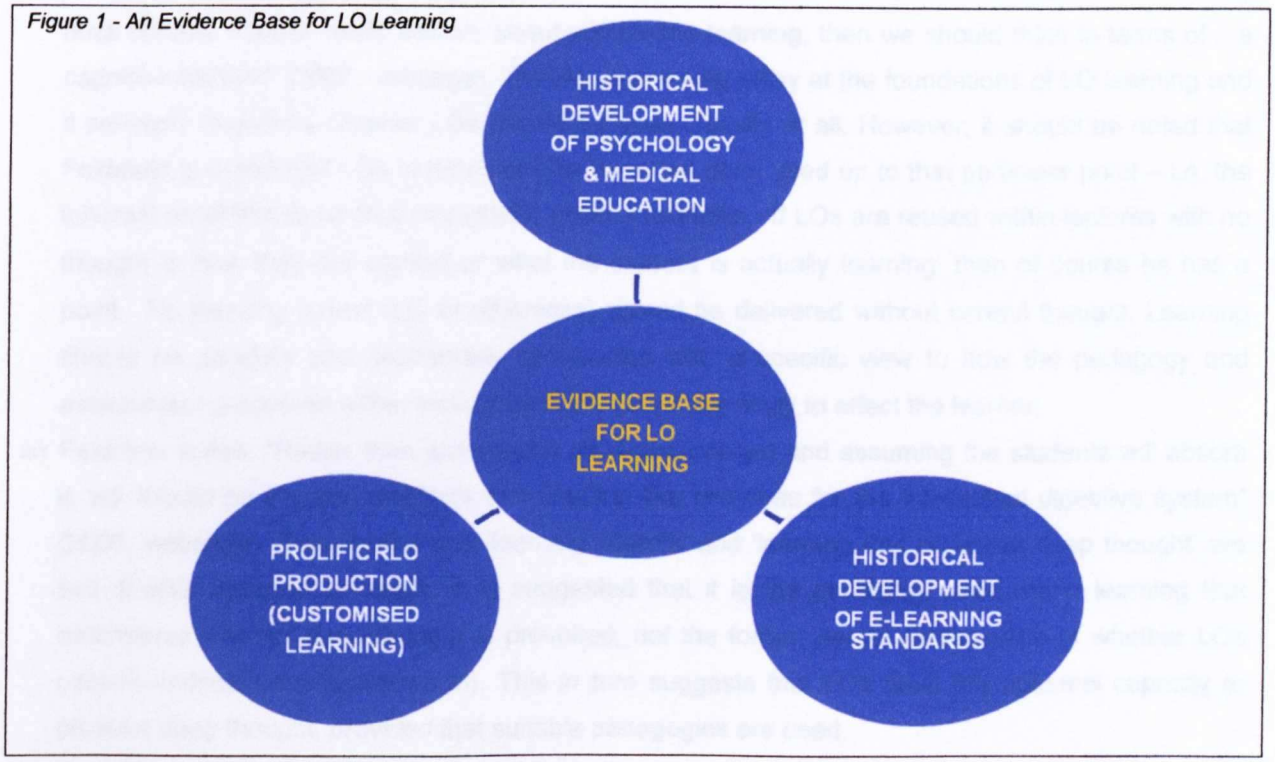
As the concept of LO learning is original, there is no evidence base for LO learning. It is presently unresearched. Problems preventing clarification appear insurmountable for some (Calbraith 2010b). I suggest that the evidence base for LO learning currently appears to lie between several different perspectives (See Figure 1):

i) Historical development of e-learning – i.e. the concept of RLOs in 2004 was seen to have the potential to remedy many web-based learning pitfalls (i.e. lack of human contact - Kruse 2002, bandwidth limitations and high fixed costs - James 2002) but a lack of national standards initially hampered progress. Development since should mean that pitfalls can now be avoided;

ii) In 2005 prolific RLO production (in response to a desire for 'customised' learning) was seen but negative aspects for widespread adoption should be noted. The Association of Learning Technology stated that 'mass learning' is conventionally 'late' as a widespread adopter, and noted that proper testing of LOs were needed to ensure quality (ALT 2005). This hints that a systematic approach may be necessary to positively establish widespread LO adoption and/or build an evidence base;

iii) Parallels with the historical development of psychology and medical education are useful when considering further development of LO Learning's evidence base. This leads to the suggestion that systematic reviews may be helpful. Provided they are carefully planned and appropriate for the issue in question, systematic reviews have been shown to have the potential to establish a robust evidence base. The negatives of this approach in an educational context are not yet fully known due to so few people attempting educational systematic reviews for obvious reasons.

Figure 1 - An Evidence Base for LO Learning



Once complete, this thesis will revisit this discussion in order to provide a more extensive description regarding where the evidence base for LO can, does, and should sit.

1.4 Why This Research? Why Now? Why Me?

Despite past excitement, mass adoption of LOs (particularly reusable ones) has since been seen to be nationally unfeasible on the scale that was originally imagined. In some part, this was due to the

smallest reusable element of the LO being the LO itself. Morales et al 2004 correct this flaw by developing GLOs (Generative Learning Objects) where the smallest divisible elements are the content which can then be used with generic LO templates. It should be noted that GLOs have also had their own criticisms: "GLOs are subjective, post hoc measures of factors only indirectly related to learning" (Brown 2007, p24). A decade on, many working with LOs are still wrestling with LO reusability. Since development time, cost effectiveness, and overall educator development are still difficult issues for RLOs, LOs are in danger of being swept along with RLO criticisms. In this light some may feel that LOs have perhaps 'gone past their peak' and have 'had their day' and feel justified in making a valid accusation. Consequently, *LOs* per se may not be taken as a serious method for today's learning environments. It is for this reason that the new conceptual emphasis I apply here, the main potential of my thesis, and pressing need for my thesis requires explanation as it may not be immediately evident to some. For example, Feldstein (2006) believes that the term 'Learning object' has actually become 'harmful'. My thesis shows that LOs are a viable learning format, so it is vital that accusations against LOs be clearly addressed without delay:-

i) It is essential to remember that my thesis' emphasis is clearly different to GLOs – i.e. the emphasis is on the *intrinsic educational worth* of LOs, *not* on reuse;
ii) Feldstein states that the term learning object "hides the same old, bad lecture model behind a sexy buzz phrase. If we're really serious about stimulating learning, then we should think in terms of... a *cognitive catalyst*" (2006, webpage). This appears to dig away at the foundations of LO learning and it seriously questions whether LOs should be used digitally at all. However, it should be noted that Feldstein is thinking of LOs in terms of how they had been used up to that particular point – i.e. the inherent emphasis is on their *reusability,* unlike this thesis. If LOs are reused within lectures with no thought to how they are applied or what the student is actually learning, then of course he has a point. No learning format (LO or otherwise) should be delivered without careful thought. Learning should be carefully and deliberately constructed with a specific view to how the pedagogy and assessment processes within evaluation approaches are likely to affect the learner;
iii) Feldstein states: "Rather than just serving up digital content and assuming the students will absorb it, we should be creating artefacts that function like enzymes for the intellectual digestive system" (2006, webpage). This implies that 'learning objects' and 'learning that provokes deep thought' are two directly opposed concepts. It is suggested that it is the *pedagogy* used within learning that determines whether deep thought is provoked, *not* the format per se (irrespective of whether LOs contain intrinsic educational worth). This in turn suggests that LOs have the potential capacity to provoke deep thought, *provided* that suitable pedagogies are used;
iv) Feldstein's phrase 'creating artefacts that function like enzymes for the intellectual digestive system' signifies one of the desires for this thesis in that it hopes to discover potential strategies, frameworks or approaches that provoke deep learning within a learning object context. Feldstein adds to this latter statement saying "without explicit thought about how the digital object in question will provoke a particular cognitive process in a learner, we're *shooting blind*". Unless a learner actually learns and at least some of the underlying mechanisms by which this occurs are uncovered (by more systematic means than 'hit/miss' or trial and error' type methodologies), it is agreed that the learning is liable to be 'misinformed' and unlikely to be learner-centred;

v) It should be noted that only three people in this field have made accusations against LOs – Downes (who has since retracted his accusations), Mayes (who has since said 'I think I was wrong'), and Feldstein (as above).

In summary, providing that i) careful thought is given to how the LOs are constructed, ii) suitable pedagogies are used and iii) explicit thought is also given as to how deep thought will be provoked in the learner, all accusations against the use of LOs can be refuted. This also goes some way to answering the 'Why now?' question regarding this research. It is precisely *because* educators in practice were (and still are to a large extent) 'shooting blind' with regard to LO learning that this research was undertaken.

So, why me? Over the years I have refereed many nursing/educational technology journals and conference papers (regarding appropriate research techniques) and feel my understanding of methodology is a positive asset. This meant that I could weigh several choices concerning methodological approaches (and their various merits) could be weighed against one another. It also means that I am not afraid to create, develop and test *new* methodologies if that is what is truly required. Hopefully my tenacious nature and genuine interest will create high quality research, and perhaps press on to persist where others have not succeeded.

1.5 The Broad Aims of This Thesis

The primary aim of this thesis focuses on the identification of *what* research, evaluation and pedagogical approaches are effective in LO practice and *why,* in order to create a rigorous basis (and hopefully theory) on which to build. The secondary aim was to ultimately understand some of the underlying pedagogical and evaluation principles at work when learning with Learning Objects. The final aim was to develop effective evaluation and pedagogical approaches that on completion of this research educators can use to develop their own work.

Although not a formal aim, it would be interesting to see if the above aims suggest how an evidence base may be built. It would also be interesting to see whether LO learning requires 'branch', 'loop' or 'branch and loop' learning systems. The scope of this thesis includes all literature and use of LOs in practice, and all adult learners' engaged in under/postgraduate university courses in the disciplines of medicine and nursing on two different sites. It is hoped, if appropriate, that findings will be generalisable to all medical, nursing and/or health populations.

1.6 How Should This Research be Done?

As previously stated, researchers at this stage had not really identified valid starting points for LO research beyond 'Here seems as good a place as any'. However, the pervading sense was that a more systematic foundation of evidence was desperately needed: "There was no systematic evaluation of the markets, no thorough and robust market research and no understanding of consumer demand (Education & Skills Committee, 2005, p17). I decided that a systematic deductive methodology could

delineate exactly 'what' was working in practice, however I noted that deductive methods often do not elucidate concise reasons as to 'why' the approaches found are effective when used alone. An explanation of *why* something is working often demands a more inductive approach. As both 'what' works and 'why' were equally important to me, two diametrically opposed strategies were obviously required. This immediately suggested a mixed method approach. I therefore take a similar view to Myers & Hasse (1988) in that the qualitative and quantitative data generated will hopefully 'supply each others lack'. Polit and Hungler (1999) describe many advantages to the mixed method approach: complementarity, enhanced theoretical insights, incrementality, enhanced validity, understanding relationships, and theory building (these will be discussed in detail together with any limitations found, in chapter 1C). They also state: "Multi-method research is often used to develop a comprehensive understanding of a construct or to validate the construct's dimensions" (p260). As constructs, mapping out construct dimensions, and theory-building are all desired this approach appeared to be appropriate. So as not bias findings, deductive literature searching will be performed *after* completion of the inductive method. No established LO research techniques currently exist - if these two opposing styles produce similar findings this may also 'triangulate' any hypotheses formed. In turn this may provide an evidence base on which to build theory.
It was anticipated that two general capabilities would be required of potential deductive methodologies: i) It should identify evaluation and pedagogical *approaches* capable of acting as 'pilot' studies to test best existing strategies, and how each applies to the target samples (i.e. doctors and nurses); and ii) the approaches found should be capable of allowing me to formulate further approaches/hypotheses which could be tested in this thesis's main research study.

There are several important questions on which this thesis seeks to gain knowledge in the course of its investigations: 'To what extent should 'non-traditional' pedagogy and evaluation be incorporated?', 'How flexible do frameworks/approaches need to be?' Scriven's (1980) approach is flexible and sees formative materials becoming summative as a new phase of evaluation begins. Sheard & Markham (2005, p354) state: "evaluation of any web-based learning evaluation must encompass not only the educational process but also the process associated with the functional usability of the technology. There is then a need to explore possible models of evaluation that allow flexibility and sensitivity to this complexity". This in turn begs further questions: Under what circumstances should approaches be flexible or sensitive to technology? What models are currently available? If any exist, how well do they perform? What is lacking? If, in the course of this thesis new methods are discovered, perhaps Postman's (1992) statement should be noted: "Technological change does not just add something, it's ecological, it changes everything". This quote highlights the importance of the LO learning *process*. This process is seen as integral to the factors that make LO learning effective.

As this thesis presents an unlikely marriage of methods, my thought processes concerning the most appropriate research methodology will now be outlined. As there was no recognised starting point for research of this kind, the first step was to find out exactly *what LO* research had been done, and specifically what was working in practice regarding LO pedagogy and evaluation. Several approaches were considered - LO evaluations, literature reviews, narrative reviews, systematic reviews, and case studies. These five options will now be discussed.

The first option was to seek out actual LO examples (nationally and internationally) and make an evaluation concerning their 'good' and 'bad' points. As few repositories and universally agreed standards existed at this point this was considered to be almost an impossible task. It was therefore quickly discounted as a possible research method on the grounds of

i) No-one knew how many studies would be available (permission, access, existence, etc);
ii) there was the distinct possibility (due to i) above) that too few LOs would be evaluated rendering the thesis unfeasible;
iii) this option would not show how learners actually used LOs, what research had been done or why it was successful. It risked overlooking the *underlying* and perhaps crucial meanings of LO pedagogy/evaluation;
iv) creation of generic principles may be possible but the only way of assessing 'adequate LO rigor' would be in relation to each other. This may not have the capability of establishing 'norm-based' criteria for LO benchmarking.

The second option was to conduct a simple literature review on LO research, pedagogy and evaluation (The term 'review' being defined as "the synthesizing of results and conclusions of two or more publications on a given topic" Sackett et al 1996, p71). However, this appeared to be inadequate – i.e. it may show *what* had been done, but may not *necessarily* show all potential outcomes or the level of rigor they had been evaluated to. This was an important omission if an evidence base is considered. I also felt that a simple literature review may not allow generalisations to the wider population due to a lack of robust baseline (as this thesis seeks to provide 'useful' pedagogies and evaluation strategies to the wider population this too was an important omission).

The third option of 'thorough literature review plus several ethnographic-type studies' seemed possible. However, these would need to be performed on many different sites, compared, tested and then presented as a 'case series'. Higgins et al (2002, p52) support this reasoning: "The case study offers some insights for a particular review, in which sub groupings of trials not specified in advance (are) presented - due to unanticipated differences in adverse effect profiles and to affirm wider applicability of the findings". This hints that this approach may be promising with regards to my research aims. It would provide an inductive/exploratory approach which could encompass learners' perspectives. Providing all work was done by me alone, researcher bias could be evaluated firsthand. Overall, this had much to commend it, but some large potential problems were evident:

i) Some may consider case studies to be 'ethnographical' by nature as they present a 'snapshot' of the question under review at a certain time in a certain context. Polit and Hungler (1999) identify the main disadvantage as a lack of generalisability hence ethnographically-based case series may not be viewed by some as 'gold standard evidence' (unless a vast amount of case studies are conducted and condensed into two or three typifying 'model' studies);
ii) The timeframe/scope of this thesis does not allow for a vast number of case studies on many different sites to be undertaken (much less their obligatory ethical approvals);
iii) As i) and ii) above indicate, a large amount of studies would have to be performed. If I was to solve this problem by enlisting lots of case study coordinators based at each site, this in itself could create problems: a) the time needed to find/train willing staff to undertake the research would be enormous. It is suggested that, given the scope of any PhD thesis, this would not have been time well spent –

i.e. busy staff would have had to undertake this research for a PhD student (unknown to them), in their own time, unpaid. It is clear that a question of motivation could be a huge problem;

iv) I may not have direct access to students – access was likely to be reliant on restrictive and possibly incompatible ethical approval (causing methodological and generalisability problems), and may not be 'firsthand';

v) If lots of subgroup analyses are performed (as warranted for this thesis), subgroups could become so dissimilar on different sites that they become unrelated thereby pulling data sources 'apart'. As I genuinely desired to pull both literature and evidence *together* (in order to make some kind of sense of it), this was seen as unhelpful;

vi) Higgins et al (2002, p60) warn: "The case study also raises the problem of pressure from peer reviewers to incorporate additional subgroup analyses". Higgins therefore sees an innate compulsion within case study methodology for coordinators to describe not just subgroups analyses in depth, but also their unique qualities. Ethnographical research would further underline this individuality. Whilst it is true strict adherence to my protocols for case studies could identify 'innate distinctiveness' (*i.e.* similar pedagogical and evaluation principles to aid assessment to some degree) it was felt that the limitations outweighed the benefits.

These potential hazards were considered a high risk strategy for a PhD – i.e. i) there was a high probability that the research would be difficult to manage, ii) it would be impossible for all sites to agree ethical considerations in a timely manner if at all and provide reasonable parity; iii) geographical locations may necessitate frequent travel to train coordinators and may therefore not be time/cost effective; iv) the research could rapidly become unwieldy with no 'quick fix' options available; v) there was no *guaranteed discovery* of pedagogical or evaluation similarities using this method; and vi) given the perceived lack of 'reliable comparison ability' – I could not be sure (even if the time was taken to match cases) that findings would be applicable to the wider population. As a result, the pragmatic decision taken was to reject the literature review method, ethnographical approach, and case series format (and all combinations of these) as possible methods.

The fourth option was to use a narrative review. Narrative reviews are often similar to systematic reviews except often no meta-analysis or statistics are involved due to the more narrative nature of the data. This option was not deemed to be appropriate due to the following reasons:

i) Cook and Mulrow (1997) state that research summaries lacking explicit descriptions of systematic methods are often called 'narrative reviews' by default. Green (2005, p271) warns that some narrative reviews may be no more than "a subjective assessment by an expert using a select group of studies to support their conclusion". I desired a format capable of more than this;

ii) If educators were to be able to use explicit pedagogical and evaluation strategies formed from this research, generic principles for use may be needed. It is suggested that narrative reviews were more likely to produce 'informed guesses' than the robust generic principles hoped for;

iii) If an evidence base was to be considered, explicit descriptors of the systematic methods were required so that others could replicate, test, verify or refute any eventual claims of this research;

iv) On talking to experienced narrative reviewers it became obvious that what I was proposing was not really in the 'true' ethos of narrative reviews;

As a result, the narrative review method was also rejected.

The fifth option was to use Grounded theory on its own using existing data. However, this too had potential flaws:

i) If a robust evidence base was to be considered and of practical use to educators, this would mean ideally that *all data in existence* should be gathered and put through the grounded theory process. As source data from published studies, conferences and other formats are so incredibly large and disparate (and no specific established or tested hypotheses is recorded in the literature), it is suggested that this would be a monumental gathering and coding task;

ii) If undertaking grounded theory in the manner outlined in i) above, there were few ways of knowing the exact potential bias with regard to the fundamental constituents of each study under examination. Those conducting the studies under question could have been asked about this but this approach relies on the honesty of the researchers concerning their own work. It is suggested that reliance on researchers' honesty may not be an adequate basis on which to build and evaluate the educational worth of LO pedagogies and evaluation approaches;

iii) There is no evidence recorded yet with regard to which types of source data are the best for LO learning within grounded theory. Even if the difficulties in i) and ii) above were conquered this would not *automatically* ensure that *all* important concepts had actually been covered.

As a result, Grounded theory used *on its own with pre-existing data* in its pure form was rejected as a viable option for this thesis (This will later be commented on as to whether this was a valid decision).

The final option considered was the Systematic review format, with or without meta-analysis. The term 'systematic review' is defined as: "a review striving to comprehensively identify and track down all the literature on a given topic" (Green 2005, p271). Some authors believe that only those looking at Randomised Controlled trials (RCTs) are *true* systematic reviews (Egger et al 2001). Some definitions include looking at observational studies also but Greenhalgh (1997, p109) expands on these saying: "A systematic review is an overview of *primary studies* which contains an explicit statement of objectives, materials, and methods and has been conducted according to explicit and reproducible methodology". This usually includes "finding, selecting, appraising, synthesising and reporting... evidence and meta-analysis for the specific statistical technique of combining the data from individual studies" (Green 2005, p272). When searching for RCTs only one education RCT was found to exist. For this reason both 'RCT only' and 'RCT & observational studies' systematic review formats were deemed to be inappropriate and a little 'previous' for this research. Greenhalgh's definition was therefore adopted.

1.7 The Chosen Inductive Method

Despite my decision to use a systematic review format for this thesis in order to provide the necessary evidence base, I recognised that this was unlikely to tell me much about *why* any LO research/ pedagogies/evaluation approaches discovered by the intended systematic reviews were effective.

Grounded theory used alone, and Grounded theory based on previous research has already been discounted as potential formats for this research. However there was the possibility that Grounded theory using *new* data would be appropriate. Advantages of this were: i) If the data was gathered by me, the full research conditions and context would be known and therefore 'controlled' by me. As such,

the full research conditions and context would be known and therefore 'controlled' by me. As such, potential bias could be prevented/limited and/or estimated firsthand; ii) Providing that steps were taken to limit and/or prevent disadvantages of the method (as previously outlined), this method would be appropriate due to the desire to render underlying pedagogies and evaluation approaches explicit. This was therefore seen to be an appropriate method to discover *why* certain pedagogies/evaluation approaches appear to be effective in practice and was therefore adopted.

I intended to perform usability studies (before grounded theory started) to ensure that technical faults in the intended LOs did not become extraneous variables later down the line (i.e. when testing pedagogies/evaluation approaches discovered from the grounded theory and systematic reviews). New source data would be obtained by observing learners during LO use, making field notes, and then using a semi-structured format (NB this appears to be at odds with classic grounded theory methodology where participants are often encouraged to talk about anything of their choosing. However, it was the intention to allow participants to talk freely about what they saw during the observations but, to ensure all gaps concerning technical aspects were covered, interviews using the usability questionnaire questions as a basis would be necessary. This said, it was intended that usability questionnaires and the Grounded Theory method would be tested out before use - see 'preliminary work' chapter. Hence a final decision on the best type of new source data would be left until after this had been completed).

1.8 The Chosen Deductive Method

Advantages of the Systematic Review Format

The systematic review approach has much to commend it:-

i) Systematic reviews establish whether findings vary significantly by particular subsets (Mulrow 1994). Greenhalgh (1997) agrees: "Reasons for heterogeneity can be identified and new hypotheses generated about particular subgroups ". This was one of the primary reasons why systematic review methodology was considered. If it could be shown that subsets (e.g. location) had similar amounts of rigor (using the same pedagogy or evaluation approach) this would be very useful in practice. It would show potential 'transferability';

ii) Systematic reviews establish whether scientific findings are consistent and can be generalised across populations and settings (Chalmers & Altman 1995). Thus it may have the power to predict learning outcome;

iii) It allows theory to be extracted from what already exists - this serves as a basis on which to estimate levels of success or effective practice. It could also form a baseline for further research;

iv) It extracts theory systematically - this highlights not only areas that could benefit from greater attention, but also those that are well researched: "A systematic review is a scientific tool that can be used to appraise, summarise, and communicate the results and implications of otherwise unmanageable quantities of research" (p271, Green 2005). Original themes may emerge after pulling the literature together in this way;

v) It lends itself well to LO/e-learning literature in that its systematic nature serves to assemble disparate but important elements: "Systematic reviews are of particular value in bringing together a

number of separately conducted studies, sometimes with conflicting findings, and synthesizing their results" (p271, Green 2005). In short, " Large amounts of information can be assimilated quickly by healthcare providers, researchers, and policymakers" (p109, Greenhalgh 1997);

vi) It extracts theory from a defined starting point. Selecting and justifying a starting point for research appears to be one of the major difficulties in LO learning. The justification 'it works in practice' may currently offer one such starting point but undoubtedly will later be seen as inadequate/an incomplete justification. (Interestingly, when conducting a preliminary literature review of LO research literature across disciplines, a large percentage of researchers had chosen review methodologies to evaluate their projects/research. Although this does not necessarily indicate that this is the best LO evaluation method it does indicate that many researchers have started from an intuitively similar premise);

vii) It minimises some of the bias encountered in small trials/projects "where results may not be robust against chance variation if the effects being investigated are small" (Green 2005, p273);

viii) Some of the explicit methods used in systematic reviews can limit bias, improve reliability and accuracy. Meta-analyses can increase power (Chalmers & Altman 1995). Hence, explicit methods limit bias in identifying and rejecting studies (Greenhalgh 1997);

ix) There is often too much data for even specialists to keep up with (Harrigan, p1154). Similarly, systematic reviews are a potential method for clinicians to overcome access and interpretation with regard to evidence to inform their practice. Green (2005, p270) state "Systematic reviews aim to inform and facilitate this process through research synthesis of multiple studies, enabling increased and efficient access to evidence";

x) A systematic review is generally the best form of evidence (Glasziou et al 2004); Conclusions are more reliable and accurate because of methods used (Greenhalgh 1997); and delay between research discoveries and implementation of effective diagnostic and therapeutic strategies may be reduced (Greenhalgh 1997). Obviously, this depends on what is being investigated;

xi) With regard to interventions, using a single study method is limited due to sampling variability. The systematic review method may reduce this if designed and executed well: "effect estimates will vary, even between studies performed in exactly the same way in identical populations so a single study often fails to detect, or exclude with certainty, a modest but important difference in the effects of two therapies. (One) trial may thus show no statistically significant treatment effect when in reality a clinically important effect exists... a false-negative result" (Egger et al 2001, p480);

xii) Finally, the systematic review format appears to be in line with emerging educational ideals. It allows the development of best evidence within education and e-learning.

Pitfalls of the Systematic Review Format

Due to its many advantages I decided that a systematic review format was possible and would be adopted but a 'more inclusive than normal' type would be required (one where all types of research could be included). Despite the clear advantages of systematic reviews for my thesis, several 'notes of caution' should be exercised when choosing appropriate methodology, e.g.:-

i) Although many aspects in education and e-learning can be measured quantitatively (e.g. length of time spent on learning activities, number of times direct questioning is used within sessions, etc),

others cannot (e.g. attitudes to learning). For this reason, the thesis will also include one-to-one interviewing to extrapolate the more personal, underlying data;

ii) Some confounding variables (e.g. unforeseen curriculum changes) could make other variables very difficult if not impossible to measure. This begs the question 'Are review formats appropriate for LO learning? Chalmers & Altman (1995) allude to this when they draw attention to the current validity of reviews, leading some authors to ask 'Are narrative reviews are a better option for e-learning?' The jury is still appears to be out but Chalmers & Altman (1995) highlight difficulties with narrative reviews also saying that it is not always possible to judge whether they are trustworthy in their present state - i.e. their objectives and methods are not often explicit. This shows a fundamental yet crucial point for *all* reviews – *they must be explicit*. For this reason the systematic reviews used in this thesis will make their objectives and methods unequivocal;

iii) Reviews in education and e-learning (in their present state) cannot reliably tell us much despite obvious benefits in science and medicine. This is due partly to lack of clarity in the way they are designed, reported and analysed. After all, systematic reviews in education are in their infancy. Begg et al (1996) recognise improving the quality of RCTs in particular is crucial. Moher et al (2006, p1503) describe the 'cognitive dissonance' in medical RCTs between "the increasing sophistication of the design ... and the apparent lack of care which they have been reported" for the first *fifty* years that they existed. Although the quality of reporting RCTs within medicine has long since been recognised and important moves towards rigorous reporting have been established by the likes of CONSORT, little work has been performed to date concerning the quality of reporting in education or e-learning studies. It is accepted that full RCTs will probably not be appropriate for most forms of educational research. However, where a high level of quantitative rigor is required, reviews may be needed. Thus I will ensure that the systematic reviews undertaken in this thesis are of high quality, and are 'transparent' in their design and the way they are reported/analysed;

iv) There are possible disadvantages when using systematic reviews as evaluation models within E-learning: "The available models for evaluation...tend to have limited availability within the development of process tools" (Sheard and Markham 2005, p360). They warn that systematic frameworks "are not necessarily readily responsive or adaptive to possible changing evaluation needs" and "in the evaluation of any web-based learning environment...the evaluation must encompass not only the educational process but also the process associated with the functional usability of the technology" (Sheard and Markham 2005, p367). This brings us neatly back to the need for a method or model that allows both technology and pedagogy to work together and underlines the strong link with practice. Sheard and Markham (2005) conclude that there is a need to explore possible models of evaluation that allow 'flexibility and sensitivity to this complexity'. Thus an emphasis on those components plus what works in practice and why will be given;

v) "Systematic reviews do not replace the need for basic research... to identify appropriate... questions and formulate promising hypotheses" (Green 2005, p273). This highlights the need for an inductive part to the thesis.

I felt that a few simple steps could be undertaken to minimise (if not eliminate) potential review pitfalls. *Potential* difficulty should not in itself lead to undertaking other methods by default, nor should it negate the necessity of doing reviews if that is what is truly required – even if this means providing slightly new methodology to cope with it. Of course, this is not to suggest that it is always necessary or even desirable. In some cases it may be more appropriate to simply wait for the field to develop further.

Having weighed all the evidence, I felt that the large number of actual advantages afforded by this method outweighed potential problems. By applying a greater level of design, data collection and reporting of research studies; and applying a greater level of consistency in terms of pedagogy, learning theory, and general research practice (necessary if a coherent and user-friendly evidence base for e-learning is to be appropriately designed and established), systematic reviews would be both appropriate and possible. As there was a clear need to evaluate LOs systematically the following will be performed:

i) analysis of non-parametric data as well as parametric data;
ii) elimination of any obvious educational confounding variables as far as possible,
iii) explicit objectives, materials and methods;
iv) clarity in the way reviews are designed, reported and analysed;
iv) 'flexibility' as far as possible to allow both the assessment processes *and* products within evaluation /pedagogy to be synthesised;
v) evaluation as to whether systematic reviews truly are the best method for this part of the thesis.

These will be assessed during the course of the systematic reviews and evaluations regarding the appropriateness of the method given throughout. Clearly, the *amount* of desired flexibility required for the subject to be studied within a review format should be considered. Methodologies once established, should then be carefully measured and tested in many different settings, institutions and disciplines to ensure rigor and generalisabilty. In this way, pedagogical/evaluation methodologies and best practice can be developed.

To my knowledge, relevant 'opposing stances' to this thesis do not exist with regard to LO learning because 'LO Learning' is a unique concept (and so little research has been done in the field of LO pedagogy and evaluation). Certainly, even at the point of writing up, no-one has yet attempted a review in this area or proposed a similar concept. As this thesis takes an original stance concerning LOs, there is the possibility that the predetermined and anticipated level of inclusion and rigor 'cut-off' for the systematic review may prove to be 'too high' when actually conducting systematic reviews (i.e. too few studies included/*all* studies excluded). There is also a real potential for the available literature to be so small that reviews would be rendered completely inappropriate. Should any of these instances prove to be the case in reality it is anticipated that other methods would be sought to correct the level, or discard the method altogether. Also it is not known whether meta-analysis or statistical analysis will be possible. If possible, they will be included together with a discussion on their potential merits and disadvantages.

1.9 The Compatibility of Chosen Deductive and Inductive Methods

The compatibility of deductive and inductive methods was considered. Polit and Hungler's (1989) many advantages were considered. As previously stated, Myers and Hasse (1988) see them as 'complementary' however there was also a potential risk of each being so specific as to what was going on and why that it may be difficult to marry the two effectively. If this is the case an evaluation will be taken post-inductive/deductive analysis (in chapter 6 - 1C) as to how to proceed. Enhanced theoretical insights are possible however there is a potential risk that triangulation may not be possible due to the nature of complementarity found. If this is the case no action will be taken as the main role for each of

the chosen methods is to uncover each others 'gaps'. However an evaluation will be made in Chapter 6 (1C) as to whether this impacts how the main research for Phase 2 will progress and any adjustments will be made at that point. Incremental progress is possible but a large degree of iteration may be needed (as this thesis is developing theory). Enhanced validity is also possible; however there may potentially be some phenomena that defy identification/explanation by *both* approaches. If this is the case understanding relationships may be difficult and an evaluation as to build theory in the most pertinent way will be made (Integration of deductive/inductive approaches will be discussed in detail together with any limitations found, in Chapter 6 - 1C). As there were only potential risks, and the advantages were in-line with the desires for this research, I decided theoretically that the mixture of grounded theory for the inductive part of this thesis and systematic reviews for the deductive part were compatible formats (as this mixture would produce different emphases on what was working and why). How well they integrated in practice and whether this was truly a good decision will be discussed in detail in Chapter 6.

1.10. What LO Theory and Instructional Design Theory (IDT) Should be Used?

The little evidence that exists regarding LO theory and IDT is largely in relation to *reuse*. Despite this paucity, it was clear to me that *some* form of instructional design should be used due to the learning's digital nature (N.B. Ideally this too should be researched alongside pedagogy, evaluation and the learning process but it was recognised that additionally researching the instructional design would be beyond the scope of this work). Therefore since patterns for instructional design have since been well documented, since educationalists have adopted IDT to aid learning when using the computer as part of the delivery system, and since this research is likely to include large amounts of data, a design capable of being 'wide' enough to allow development of LO learning whilst simultaneously being 'narrow' enough to hold the process together was required. I personally believe that the underlying IDT should not dictate what should be included (it should not force component elements into certain 'moulds', nor should it limit LOs to one form of e-learning, nor allow the research to become compromised or unfocused by its addition). Instead it should structure the process and hold it together. As such, only two IDTs held appropriate attributes – i) Gagné's Instructional Design Theory (2004); and ii) The Instructional Design Institute's model - adapted from K.L Gustafson's model (2003). Gagne's design was considered first (Figure 2). In brief, Gagné's IDT provides an analysis of the learning to be accomplished (1-6), then translates this into a design for instructional events which prompts/supports the internal processes of the learner (7-9). These are then tested, used and evaluated (10-11) (Petry, Mouton and Reigluth 1987).

Gagné's design is based on different types of learning outcome. Each of the nine basic instructional events have variations for the type of learning outcome and require different learning activities (and therefore different instructional conditions). It is suggested that developing instruction involves analyzing requirements, selecting media and designing the instructional events. However, little is offered concerning the process of creating the instructional materials themselves. As I felt that a clear analysis of the instructional process would 'tease out' some of the underlying reasons for why certain

pedagogies and evaluation strategies worked, this was felt to be an important omission. Gagné's IDT was therefore rejected.

Figure 2: Gagné's Instructional Design Theory (2004)

Analyzing requirements for learning works back from the intended learning goal

1. Identify types of learning outcomes we wish to achieve.
2. Most learning outcomes are not simple - each outcome must be broken down into a hierarchy of dependent learning outcomes and pre-requirements to give a learning hierarchy of simple outcomes
3. Identify what conditions/processes internal to the learner must occur to achieve those outcomes.
4. Specify what external conditions/instruction must occur to achieve these internal conditions.

Selecting Media

5. Record learning context.
6. Record learner characteristics.
7. Select media for instruction - Books, whiteboard, Computer Assisted Instruction and video are common examples.

Design Instruction - planning instructional events to support learning activities

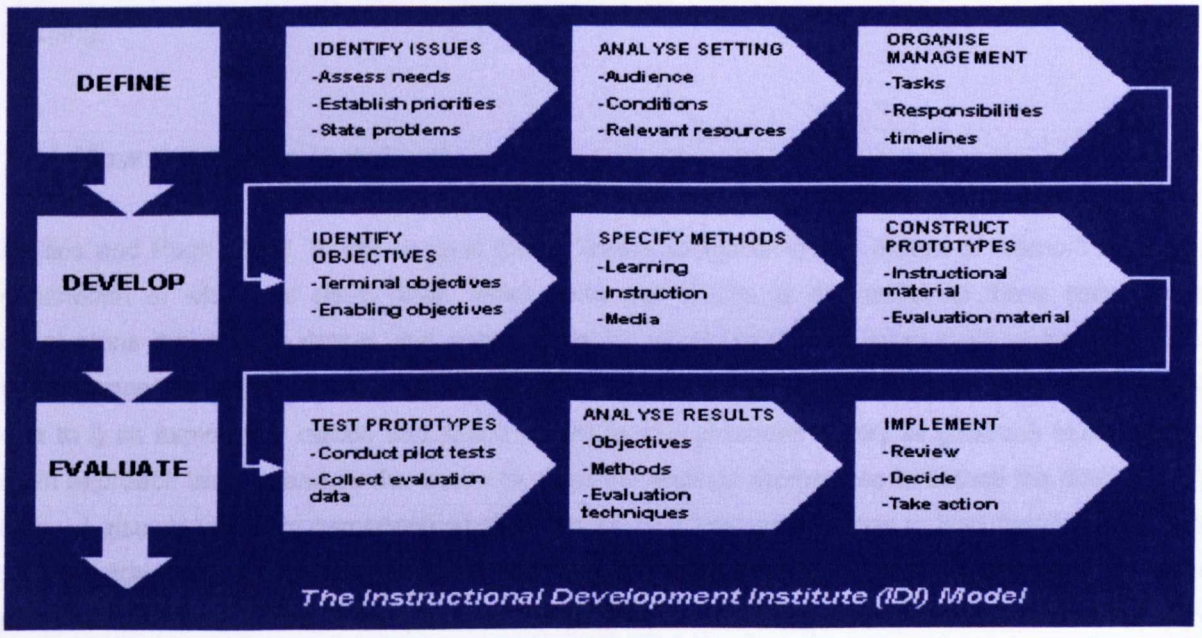
8. Plan to motivate learners by incentives, task mastery or achievements.
9. For each of the planned learning outcomes in the learning hierarchy, the Nine Instructional Events are designed relevant to the type of learning outcomes required, in the order of pre-requirements in the learning hierarchy, and with appropriate media and use of tutors.

Testing

10. Although instruction is apparently ready to use, in practice they are tested in trials with learners (formative evaluation).
11. After the instruction has been used, a summative evaluation can judge its effectiveness.

In 2003, International Skills conducted an 'Instructional Development Models' survey at Syracuse University, resulting in the IDI model adapted from Gustafson's model (Figure 3).

Figure 3: IDI Model



In many ways it is similar to Gagne's design in that it has desired outcomes. However, the different order of similar components in the IDI model places different levels of emphasis on the components. It has already been alluded to that considering and evaluating the learning process alongside the pedagogy and evaluation may be crucial. The IDI model allows in depth assessment of the learning context ('Define' stage) before consideration of objectives ('Develop' stage). Admittedly, an idea of the desired outcome has to be 'held' throughout the process - without it results in little focus, nebulous outcomes, and difficult evaluation outside of formative formats. Nevertheless, due to the fact that the IDI model allows in depth assessment of the learning context *before* consideration of objectives (and this

fitted with the inductive parts of the thesis) it was considered to be the model of choice as this permitted simultaneous evaluation of process and approaches. Once complete, this thesis will return to this model to evaluate whether this is still considered to be an appropriate model to use.

Learning Skills (2003) highlight the second reason why this was chosen. They believe that the model's basic strength is its three levels permitting 'simple initial presentation' to non-developers. This can then be developed as knowledge increases. In a nutshell, this approach had the capability of great flexibility for educators (one of its major claims is that this approach can be elaborated on). Also, an easy-to-use format was desired to help educators utilise any products (approaches/models) produced by this research in their own contexts.

Thirdly, a greater depth of *overall* testing is incorporated into the IDI model (e.g. 'Evaluate' stage) when compared to Gagne's design (parts 10 & 11). Greater guidance as to what form this should take is also given (e.g. prototype testing, analysis, and implementation). Incidentally, the stages of the IDI model have obvious parallels to PhD requirements and intended structure (i.e. PhDs in general need to provide description and analysis of the research process taken, and the design, development and testing of theory). Although this is not a major factor regarding the IDT choice, the overall design is complementary.

In short, due to the fact that it places i) less emphasis on outcome and more on process; ii) encourages a greater depth of testing throughout the process; iii) appears to be easy for non-developers to use; and iv) fits with the PhD structure, the IDI model was chosen to be used as a basis for this thesis' research. Of course, use of this model will later be reviewed concerning the fundamental premise of IDT within LO learning.

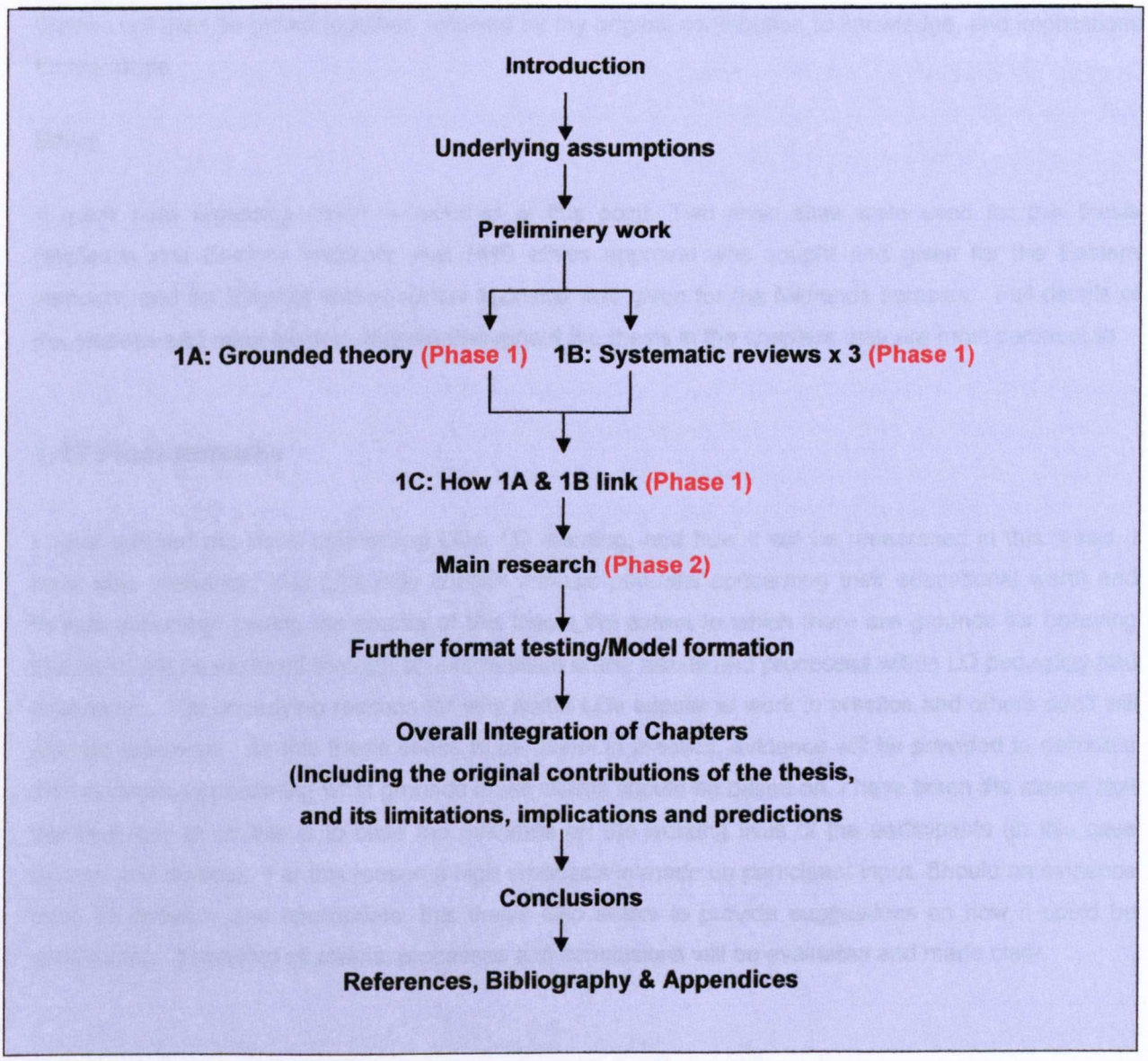
1.11 How the thesis is set out

Phillips and Pugh (2000, p. 65) suggest that a "thesis should contain a review of relevant literature, a description of what has been done, what came out of this, a discussion of these results, some conclusions that can be drawn, and suggestions for future work". The exact methods/structure of my thesis cannot be outlined in advance (beyond the choosing of grounded theory and systematic reviews) due to i) its exploratory nature and ii) the desire to use grounded theory to generate hypotheses for main approach testing; and iii) the desire to allow the findings themselves to dictate the direction to be taken. I also do not want preconceived ideas concerning best approaches to bias development hence as each stage of the research is analysed this in turn this will suggest/dictate what research method could/should be used for the next phase of development. It is therefore not possible (or advisable) to do a full background/literature review before gathering grounded theory from new source data.

For ease of reading this thesis will present the research according to the chronological development of each phase and outline what it was trying to achieve – i.e. set out chronologically with regard to sequence, time and development. As such the rationale, creation of the research method, data collection, results, discussion, analysis (and/or statistics if applicable) will be included in each section for each phase of the research (rather than being set out in the usual manner which would require repeated headings). Not only would this save an unnecessary replication of words, but more importantly I do not want to divide the research stages into constituent parts as it is important for the reader to understand

the process/chronological development made. Thus each stage states its aims/intentions, undertakes the work, outlines results, reviews the methodology used and discusses the issues. Each stage is written in future, present then past tense to give the reader a 'firsthand experience' of the developing research. It is presented as shown in the contents list to show the chronological flow of the research and how each stage suggested the next development stage (See Figure 4). In short, the unique approach taken by this thesis is therefore by design. In order for the initial grounded theory part of the research not to be biased by a literature review (which is necessary for the systematic review protocol), and in an effort to be consistent with the grounded theory method (Glaser 1998, p68), I decided that a data source for the grounded theory should be decided upon and conducted before writing the systematic review protocol. (Ideally should be done sequentially, however a slight overlap between the grounded theory and the systematic review data collection was necessary. Although this was not ideal, it should be noted that all 'source data' research for the grounded theory had been collected and coding completed before finishing the protocol and conducting systematic reviews).

Figure 4: How the Thesis Chapters Are Set Out



Both general and systematic review literature were then added to the theory in keeping with the constant comparative method. As I was able to keep methodological integrity, this unanticipated

problem is not believed to have affected the research inordinately, but it will be discussed in chapter 1A if necessary). As this thesis' findings are chronological, the 'literature that helped develop the theory' will also be chronological i.e. integral literature elements will be woven into the grounded theory. As previously stated, it was hoped that grounded theory would be an appropriate and *direct* method to discover *why* certain pedagogies and evaluation approaches appear to be effective in practice. As the intricacies of grounded theory and systematic review approaches concerning LOs in particular may be unfamiliar to some readers, a detailed description of each will be given in each corresponding chapter to orientate the reader to the methodologies' rationale.

Prior to the inductive and deductive chapters, 'scoping surveys' will be performed to establish what learners want and why. Together with a 'testing' of techniques/populations, these surveys will form the preliminary work chapter. Thus usability studies and grounded theory will form Chapter 4 (Phase 1A), and systematic reviews chapter 5 (Phase 1B). Chapter 6 will compare the findings which will dictate how the hypotheses gained should be tested. Testing will form Chapter 7 (Phase 2). If any governing principles are found, these will be incorporated into the following chapter. A discussion of all thesis themes will then be pulled together, followed by my original contribution to knowledge, and implications for the future.

Ethics

A quick note regarding ethics is required at this point. Two main sites were used for this thesis (Midlands and Eastern England). Full NHS ethics approval was sought and given for the Eastern samples, and full medical school ethical approval was given for the Midlands samples. Full details of the process and rationale will be given throughout the thesis in the chapters they are most pertinent to.

1.12 Final remarks

I have outlined my ideas concerning LOs, LO learning, and how it will be researched in this thesis. I have also postulated that LOs may contain *intrinsic potential* concerning their educational worth and flexible capability. During the course of this thesis the extent to which there are grounds for believing this claim will be explored through an examination of the issues and processes within LO pedagogy and evaluation. The underlying reasons for why some LOs appear to work in practice and others don't will also be examined. As this thesis seeks to be *useful* in practice, evidence will be provided to delineate the boundaries concerning what grounds these claims should be based on. I have taken the stance that the best way to do this is to base the evidence on the working lives of the participants (in this case doctors and nurses). For this reason a high emphasis is made on participant input. Should an evidence base be possible and appropriate, this thesis also seeks to provide suggestions on how it could be constructed. Therefore all claims, processes and conclusions will be evaluated and made clear.

* In this chapter there are some references to 'evaluation approaches'. This term has been chosen by the author to refer specifically to 'built-in evaluation processes' or 'inherent processes' within LOs. The approaches, when constructed, can of course be used for assessment.

2. UNDERLYING ASSUMPTIONS

Despite the best will in the world, researchers do not (and cannot by nature) approach subjects completely free from preconceptions, underlying assumptions, beliefs, thoughts and opinions. No researcher works in a vacuum and must bring their own subjective thoughts to the research either by default or definition. I have been interested in LOs since their inception and therefore could not help forming ideas about this area. It would be wrong for me to assume that because I *aim* to be 'bias free' that my underlying assumptions will not influence the intended research. NB The following underlying assumptions and beliefs outlined *are NOT themes and approaches that I have chosen to shape the thesis*, but are themes that I am aware may influence the interpretation of thesis findings. As the initial part of the thesis is inductive it is important to be honest about them at the outset. I have therefore set out my preconceptions to allow proper public scrutiny.

My assumptions will be described in this section and re-examined in the overall integration chapter (Chapter 9) upon completion of the research to ascertain the extent to which they have influenced the research and/or whether they have introduced bias. As the initial stage of my thesis is inductive, no background literature will be sought at this point. Instead, a literature search will be conducted before the systematic reviews, and henceforth any necessary literature will take the form as dictated by the research findings (e.g. either to start of each phase/chronological research section, for ease of reading, to set the scene, or to highlight the areas of importance).

2.1 My Underlying Assumptions, Beliefs, Thoughts and Opinions

Learning Object Theory and Instructional Design Approach - I believe that a clear analysis of the instructional process during discussion of the research results will help make clear and 'tease out' some of the underlying reasons for why certain pedagogies and evaluation strategies work. This could be seen as a positive bias towards the IDI model (Figure 3). Playing devil's advocate, if this thesis was *primarily* concerned with Instructional Design Theory (IDT) this could be a major flaw. However, as this thesis is only concerned with IDT *indirectly* this was not seen to be a major problem at this point. Of course, an evaluation of this assumption is required post-research to weigh up whether it became an actual bias in the course of the research.

Learning theory Approach - In the initial stages of this research the inductive portion will bring in learning theories only if the research dictates their necessity. It is therefore not known at this point which approaches, if any, will be most appropriate. In the deductive chapter, a review of research, pedagogies and evaluation approaches will *uncover* 'effective' existing learning theories – therefore these also are unknown at this stage. I am not aware of having any preconceived assumptions regarding either of these areas.

Theoretical Framework - No theoretical framework presently exists for building rigorously tested LO pedagogy/evaluation approaches, but this thesis may *inadvertently* provide one. If time allows, it may also create and develop usable theory. I am not aware of having any preconceived assumptions

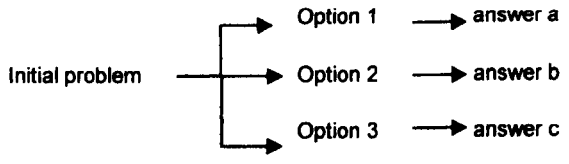
regarding this area, except perhaps that a usable theory on LO learning may be possible.

Conceptual Framework - I have an underlying belief concerning appropriate conceptual frameworks and this is closely tied into adult learning theories. My own beliefs are similar to Brookfield's approach. Brookfield (1987) discusses 'conditions that influence adult learning' and believes that these conditions are most effective when learners i) know what they might gain from the effort of learning; ii) are given the Module Guide Aims/Learning Objects; and iii) when they determine the course and pace of their own learning. All these aspects complement current learning trends – i.e. learning should be relevant to context and learner and provide a wide degree of learner choice. Brookfield believes that when the above conditions are put in place i) Learners perceive that learning is related to their own experience (Giving them ownership, deep learning, etc.); ii) The topics used are those which help them deal more effectively with their everyday problems (i.e. must be practical/relate to practice); iii) Topics relate to actual tasks and problems (i.e. direct practical applications/simulations); iv) Learning is seen to enhance job satisfaction and self-esteem (i.e. 'Lifelong Learning' and 'ongoing personal research is always needed' mentality); v) Learning incorporates elements of challenge to promote critical analysis (i.e. takes them beyond what they already know - e.g. the transition from information to knowledge, or a little knowledge to greater knowledge); and vi) Learning takes account of the needs of the organisation and society as well as development of the individual.

These 'effects of good learning conditions' will be compared with the research once completed, and will be discussed in detail with regard to whether they have either positively or negatively influenced the findings.

Learning systems approach - Due to my experience of writing LOs, I have an underlying belief concerning LO learning systems. Downey (2003) debates the various merits and disadvantages of learning materials suggesting that they have two distinct categories - 'branch' or 'loop' learning systems. From experience, I know that learners want to choose what they learn in order to make it highly relevant (hence the need for different branches of learning - Figure 5):

Figure 5: Branch Learning



In addition, they often want to revisit parts of the learning (hence the need for loops – Figure 6). Thus from experience I have developed an original underlying assumption that LO learning requires *both* types of learning system when used on-line. I therefore assert that a new mixture of both 'branch' *and* 'loop' learning systems - where the branches ultimately all connect back onto it forming a loop - is possible (Figure 7). In order to shed more light on these beliefs and assumptions, the LOs will need to be designed so that the learner can choose their own route through the learning material. This was not

perceived to be a potential problem or undue bias for the research; however both this concept and its underlying assumptions will be revisited and discussed once the research is complete.

Figure 6: Loop Learning

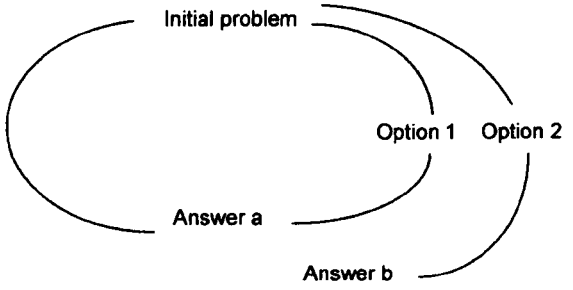
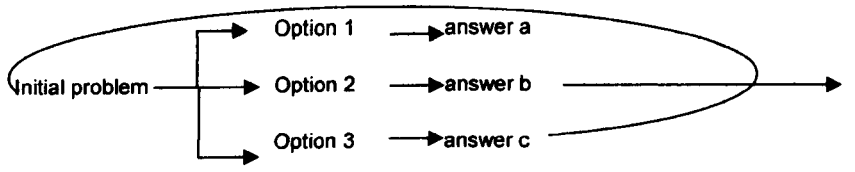
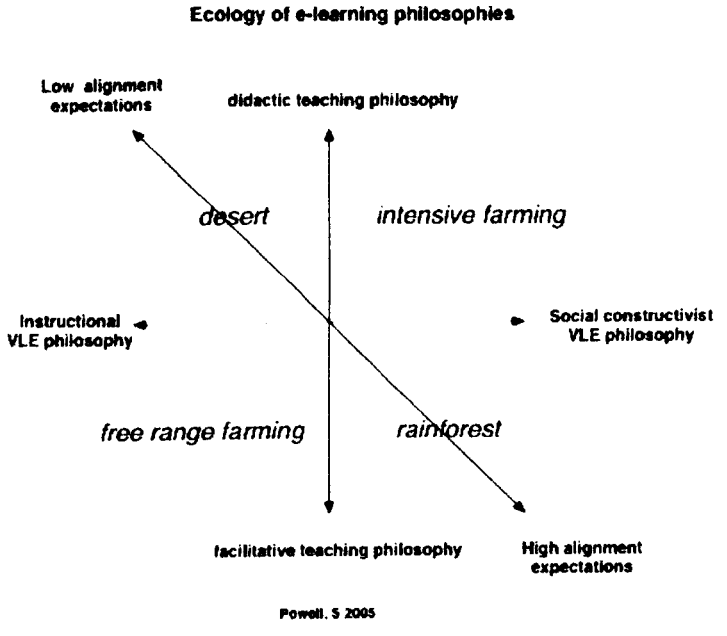


Figure 7: Branch and Loop Learning



Philosophical approach - As the thesis' starting point is inductive, it is not yet known what the underlying philosophy/philosophies may be. Traditionally, instructional e-learning material and didactic teaching methods have met with 'low alignment expectations' (Powell 2005) and it is easy to see how some methods (e.g. scanned copies of lecture notes) will not optimally fulfil either the digital learning or teaching aims. Therefore the philosophical standpoint has to be one that encompasses all the required elements outlined so far, but not rigidly dictate how this should be achieved. I have an underlying assumption that Powell's (2005) Ecology of e-learning philosophies may be helpful (Figure 8) and the extent to which this has shaped, biased or positively influenced my thesis will be evaluated later.

Figure 8: Powell's (2005) Ecology of e-learning philosophies



3. PRELIMINARY WORK

This chapter outlines how I discovered nursing/medical student preferences concerning 'which LOs are most desired' and 'why' (Part 1: needs-based surveys). This provided background data for both medicine and nursing populations, and afforded me the opportunity to test how effective my proposed data collection methods were. This chapter culminates with how and why the Usability Study format and Grounded theory from usability observations (Part 2) were decided upon for this thesis for Chapter 4 (Phase 1A). The result is a comprehensive testing of proposed thesis methods, and a plotted 'geography of the landscape'.

3.1 Survey Background

Denscombe (1998) describes surveys as 'snapshots' representing situations at a given point of time. I wanted to gather and understand existing aspects of learning preferences in my proposed populations: "surveys... are well suited to descriptive studies, but can also be used to explore aspects of a situation, or to seek explanation and provide data for testing hypotheses" (Kelley 2003, p262). I also wanted to explore what types of LOs were perceived to be 'most needed' so that LOs created later could be appropriate and relevant. Furthermore, I wanted to delineate this topic's 'boundaries' if possible: "Descriptive studies are used to estimate specific parameters in a population... and to describe associations" (Kelley 2003, p262).

Kelley (2003) propounds several survey advantages. Providing the process is well executed these include:

i) empirical data – collected and based on real-world observations. This coincides with the underlying empirical desires for this thesis;
ii) more generalisability to a population (i.e. the more people asked the greater chance the data is based on a representative sample. This is then more likely to be generalisable to a greater population). One desire for this research is to explore whether generic principles are possible/exist. Generalisability is an important aspect if principles can be appropriately applied; and
iii) large amounts of data can be produced in a short time span - as this was intended to be quick 'scoping' studies to determine 'what LOs should incorporate' I did not want to spend large amounts of time on them but nevertheless wanted comprehensive gathering of 'authentic' data.

Disadvantages include;

i) inadequate coverage of data implications. Exclusion of relevant issues/problems surrounding data can undermine the significance, so these will be discussed in each survey section;
ii) data is likely to lack details on the investigated topic – participants will therefore not just be asked for 'yes/no' answers but will be asked to volunteer reasons for their answers (i.e. open and closed questions). Generic questions will be used for all participants on all sites (See Appendix 1); and
iii) response rates are unpredictable – whether an adequate response rate will be achieved is unknown. If this becomes a problem, survey response time will either be extended or further ethics

approval sought for extending means of response. Each survey will remain 'active' for a 4-week period for participants to access/perform the survey.

It occurred to me that although Kaufman's method was intended for usability testing, I could use it in an original way to shape survey priorities, thus affording easy result translation from survey to usability study and consistency between methods. Kaufman (2006, p2) states: "You need to decide *what* you're going to test (ask). The best way to do this is to... choose features that are new, frequently used, considered troublesome or especially important. After choosing... prioritize them and write task scenarios based on them". The main research question for each survey will therefore be: 'Which area/s do you feel would be most beneficial to you that are not currently covered in your curriculum?' Efforts will be made to ensure that extraneous variables (e.g. lack of computer access/skills) do not skew results, and the order and nature of questions do not bias other answers (by considering the position of questions in relation to one another). To get an amalgamated 'profile' of the population, the surveys will be open to *all* members of the targeted populations.

To achieve a good response rate, the communication methods that each discipline is most familiar with will be used to contact them. I did not want to 'interfere' with the intended samples' normal expectations of their course. This initially suggested an Ethnographical approach (which would provide rich data from the proposed sites for the main research) with a logical development onto case studies. However, a full ethnographical approach may have worked *against* one important hope for this thesis – i.e. Exploring *similar* (not unique) elements in each site with which to form 'generic principles' if appropriate.
Secondly, I did not want the 'fullest data' possible on the subject, but 'theoretically relevant' data.
Thirdly, I had initial reservations regarding case studies as they may prove harder to evaluate due to the high possibility that cases may be extremely heterogeneous. Despite Wolcott's (1994, p94) assertion: "there is always the possibility for generalisation and often the readers themselves can make that leap" I was not reassured. Wolcott (1994) also believes that researchers can make comparisons with the other specific situations similar to the case studied and achieve 'typicality'. However, I felt that it was almost certain that comparisons of more than two sites would be necessary to achieve this. Whilst acknowledging the validity of this approach, I decided that due to i) time constraints; ii) potential difficulties in gaining an appropriate level of access permissions to many different sites; and iii) possible ethical permission 'mismatches', that a usability/grounded theory approach would serve this thesis better by i) limiting the necessary testing sites to two (therefore requiring a shorter timeframe), ii) reporting participant experience, and iii) using usability studies to provide rich data (not only to explore the learning environment but also how technology aids/interferes with the process).
Fourthly, it must not be forgotten that this was first and foremost a 'scoping' survey. This provides another valid reason against using a full ethnographical approach in that it simply was not warranted.

3.2. PART 1: Needs-Based Surveys

Part 1 outlines how the needs-based surveys were conducted. Part 2 outlines the usability studies, observation, interviews and reasons for grounded theory testing. All methods in the following two surveys (based at a Midlands-based University) will be presented here in terms of rationale, creation, data, results, discussion, and analysis.

3.2.1 Medical Student Survey (Touchstone)

Introduction, Rationale and Creation

I wanted to elicit medical students' desired LO topics and underlying perceived reasoning. I decided that students should be asked which subjects they found most difficult to learn, how they learned best, which areas would benefit most regarding the provision of new packages, what subjects/skills they most enjoyed, the reasons for these perceptions, and also questions to ascertain whether students had easy internet access. It was crucial that lack of access/computer use knowledge should not hinder the research by later becoming confounding variables. A survey for the medics was created from the generic questions I had identified (Appendix 2)

To compare results across disciplines i.e. Nursing and Medicine (and location during a later stage) I required access to all Midlands-based medical students aged between 18 and 58 years old. The total intended target population included all medical students within 5 different curricula located across 3 different sites (i.e. years 1-5, n=300). Ethics permission was sought and granted for this sample despite these students being 'over-saturated' with surveys. Given that a 'good' response rate for face-to-face surveys is 60% (and online surveys are generally less), the estimated maximum number of respondents was 180. This was deemed to be a sufficient population to build and test a valid and robust set of LOs.

The communication method most familiar to these students was used to achieve the highest possible response rates. This population usually receive all messages via their Virtual Learning Environment (VLE) and complete all surveys using Touchstone - hence 'Touchstone via VLE' was chosen as the survey delivery mode. All eligible students were invited to participate via VLE and were given 2 weeks notice concerning the forthcoming on-line survey. They were then given 4 weeks in which to access and complete the consent form, information sheet and survey 'live' on-line.

Touchstone specifications - Touchstone was originally designed as an 'in-house' web-based system and used for summative assessments (HEA 2006) but is now used for surveys. It has 3 main front pages: 'Permissions', 'Survey creation' and 'Survey store' pages. The *Permissions page* was set up with a 4-week restricted access with specified dates for students to complete the survey. Generic questions were entered into Touchstone on the *Survey creation page* for each year group producing five surveys (with specific year group questions added. Appendix 2 shows this during creation and Appendix 3 shows this after saving). Completed surveys were stored in the *Survey store* in the form students see them (Appendix 4).

Data Collection

Responses were collected automatically by Touchstone at the end of the 4-week period (An example of the survey's raw data is given in Appendix 5. A 'snapshot' of collated data is shown in Appendix 6). Due to the automatic breakdown of percentages it was easy to pick out the top student choices for LO development.

Results

A total of 105 medics responded (corresponding to 35% of the total population and 58.3% of the maximum expected response rate). As this was an 'oversaturated' population this was deemed to be a good response. 97% of respondents had easy access to the internet and 100% had access to the LOs. Anatomy, physiology, histology, and infective processes/conditions of the respiratory system came top concerning 'most needed' and 'difficult subject' lists (a breakdown of these and common subjects for all years are given in Appendix 7 and 8). Application of clinical skills was top of the 'most enjoyable' list. I created 7 computer-based LOs to include the top 5 for all year groups (i.e. Pathophysiology, Clinical examination skills, Normal Structure and function, Infective/inflammatory/metabolic processes and Respiratory systems). These top 5 were chosen to shape the design and content for the intended LOs.

Discussion and Analysis

Specific parameters of the population concerning desired topics were ascertained, and as an equal chance was given to all students to respond anonymously the data was considered to be a sufficiently valid representation of respondents needs. The decision to use a tailor-made 'in-house' questionnaire in this population was therefore valid. (NB Should these results be added to later results to form a basis to generalize findings it must be remembered that despite the response rate being good for that institution it is representative of only 21% of the overall population). In-depth information was gained concerning student learning therefore known survey method disadvantages did not adversely affect the result. Efforts made to ensure that extraneous variables (e.g. lack of computer skill/use) did not confound results seemed adequate as 97% of students had good internet connection, computer access and level of computer skills. These aspects were therefore not assumed to have unduly affected response or answers given.

Participants were asked for reasons for their answers. Initially, the Ethics Committee did not understand the importance of this and felt that it was an unnecessary inconvenience for participants. However I defended, explained and pursued this approach and approval was finally given. This proved later to be important as these reasons formed specific answers concerning i) the parameters found in the population; ii) the prevalence of similar answers; and iii) went some way towards uncovering associations between answers when tested later in different populations. This approach proved popular with busy medics as the survey took approximately 5 minutes to complete. They felt their views were being listened to, were excited about having LO input, and felt that 'special' LOs were being created as a direct result of their participation. In short, despite the 'lower-than-hoped' response rate, Touchstone provided a comprehensive 'geography of the landscape' and specific parameters of the population.

3.2.2. Nursing Student Survey (EBP)

Introduction, Rationale and Creation

I wanted to elicit nursing students' desired LO topics and underlying perceived reasoning. As before, the survey questionnaire was drawn up from generic questions (Appendix 2).

In order to later compare results across disciplines (i.e. Nursing and Medicine) and location, I required access to all nursing students at the same Midlands-based University aged between 20 and 58. The total target population was approximately 300. Given that a 'good' response rate for face-to-face surveys is 60% the estimated maximum number of probable respondents was 180. These students were also 'over-saturated' with surveys, hence only access to Evidence-based Practice (EBP) students was granted (90 nurses, i.e. 30% of the total population). This was hugely disappointing. Ramifications of this limitation are not yet known but will be revisited later.

These nurses were used to receiving paper surveys in person via their tutors at the start and end of each module. Hence paper-based/tutor contact was chosen for survey delivery as it complied with their normal practice. I identified three different EBP nursing student courses (each 8 weeks long) that fell within the research timetable. All eligible/potential participants were invited to participate, and the survey's rationale and consent forms were given out at the start each EBP course by their tutor. Each cohort were told via their tutor that they had 4 weeks to participate should they wish to, and that any completed forms would be collected at the end of this period.

Data Collection

Surveys were given out to participants as per above plan. I was not allowed personal access to the students but this was not felt to be a problem despite being totally reliant on tutors to deliver/collect surveys within the allocated timeframe. The importance of the strict timeframe and implementation details were explained them. On completion of the 4 week 'survey-active' period, 30 respondents from 2 groups had completed surveys - tutors returned these to me. The final group's tutor told me at the end of the 4-week period that she had some surveys but did not return them. A while later she returned 10 completed surveys but admitted that some participants had had as much as 8 weeks to complete them. Longer time may have elicited a more in-depth participant response and as a result no direct comparison can be made between this group and the medical students' survey. As there was no way of telling how many of these 10 had been non-compliant/contaminated all 10 were discarded.

Results

These 30 'usable' responses corresponded to 33% of the total EBP population and 55.5% of the expected maximum response rate and were deemed to be a good response due to its level of 'oversaturation'. 97% of respondents had easy access to the internet and 100% had access to computers. Similar subject areas came top concerning 'most needed' and 'difficult subject' lists (a breakdown of these and common subjects for all years are given in Appendix 9 and 10). Using new clinical skills was top of their 'most enjoyable' list. The 7 computer-based LOs created therefore fitted well with these findings so only a couple of terminology type changes were required (E.g. 'conducting portion'). Again, the top 5 subjects were chosen to shape the design and content for the intended LOs. (Unfortunately as there was no access to demographics no similatities/differences between year groups could be made).

Although the response rate was acceptable (33%) 30 surveys represented only 10% of the total nursing population and 16.6% of the expected maximum response rate. Perceptions of 30 nurses was more than adequate to provide enough information for the LOs, however this could be a major problem later if it became important that this population was representative of the total population (Ideally other cohorts at another university would have been sought and data collection recommenced at this point, however both time and departmental politics prevented this). Consequently, the results presented here represent only an *approximation of nursing student desires concerning LO topics*. If this survey had been a major part of the research the restricted access would become a major problem and the use of an alternative participant group would be vital. However, an alternate participant group was impossible at the Midlands site. Also i) as this survey was only a scoping survey; ii) the main emphasis was on preferences (rather than level/type of discipline); and iii) it had been extremely difficult to get this limited permission (producing a strong belief that further efforts to secure alternative samples would be unfruitful), I had to accept the sample gained (as limited data was better than none).

In my experience, EBP courses generally contain a broad cross-section of nurses from different specialisms and it was hoped that this was the case in the participant group. However, since no access to the whole student population was granted, I had no way of evaluating this. Estabrookes et al (2003) undertook a large study where nurses using the internet to gather EBP information were found to be representative of the whole population. Although this seems to support my experience, it would be naive to make the same link here. Firstly, Estabrookes' study was conducted in Canada and may not be representative of the UK population. Secondly, inadequate reporting means that Estabrookes study cannot reliably support or refute my experience. Whilst the EBP sample may have been adequate for this survey, demographics will need closer inspection in order to tease out any potentially misleading findings. On the positive side, these surveys were able to provide a good 'geography of the landscape' and answer the research question. The data also represents the first of its kind on nurse-based EBP LO learning. Admittedly, the medical 'geography' was much better defined which I attributed to Medicine's better 'student/computer' communication procedures.

3.2.3. Comparisons Between Needs-Based Surveys/Discussion

Both surveys had similar response rates therefore approximately 34% is anticipated for the main study (at the same university). As 4 weeks proved to be long enough for those who wanted to participate to respond, this time period was considered sufficient for the main study to gain useful data:

Table 1: Comparisons Between Needs-Based Surveys

	Mean survey age (yrs)	**Target population vs. actual population response rate**	**Data collection period (weeks)**
1a: medical students	20	300 vs. 105 (i.e. 35%)	4
1c: nursing students (EBP)	34	90 vs. 30 (i.e. 33%).	4 (those having up to 8 weeks were removed
Overall Mean	27	34% (roughly 3:1)	4

It was previously anticipated that there may be issues of survey parity due to different communication used. The response rate was very similar in each case (33%, 35%) which indicates that the differences (and therefore influences) are probably negligible. This approach will therefore be used in the main study (i.e. approaching participants using their most familiar communication method). Should it become clear that parity or reliability is suffering, this will be more extensively explored.

While nurses' Internet use at home increased over the data collection period (and was comparable with other groups), Internet use at *work* for EBP nurses was low compared with other groups despite adequate workplace access. From survey answers it was clear that the nurses valued interpersonal contact more which created a tendency to prefer personal experience and communication with colleagues/patients to on-line and traditional sources of practice knowledge. Previous work in this field has examined whether nurses have experience in using the Internet, and if so, how they use it. However, none of these studies (Lawton et al 2001; Cobb & Barid 1999; Dumas et al 2001) offer any information to explain why this particular EBP sample may prefer personal contact. Bachman & Panzarine (1998) found that increased computer use opportunities equated with increased LO value. This may indicate that these EBP students did not have enough opportunity to use (and therefore value) computers/LO throughout their course. If so, this extraneous variable issue may need closer inspection later if it is shown to skew results. Sample demographics would also need closer inspection. For this reason, and the possibility that inferences from this sample may be needed to later generalize findings to the wider population, the Midlands ethics committee was approached concerning access to demographics. Permission was denied. As such, the EBP results presented here can only represent an *approximation of nursing students' desires* concerning LO topics (the level to which they are fully representative of the whole Midlands-based nursing population cannot be verified). Using this sample in its present state as if it were representative would undoubtedly lead to a Type I error. Added to this, revisiting this population for the main research as intended was impossible due to 'faculty difficulties'. Thus despite proving to be an adequate test of intended methods, a different university will therefore be used to repeat this nursing survey (and for the main research) due to little choice in the matter. As demographical details were freely given by the students using Touchstone, there were no such difficulties for the medics.

The original aim for the main research testing was to have multi-site multi-disciplinary research with both nursing and medicine on each site – i.e. Midlands and Eastern regions - so a direct comparison between locations/disciplines could be made. An Eastern ethical committee was approached to fresh access nurses/medics. However, they too could not understand why I needed this access (despite copious explanations in person including ethics permission given for the Midlands-based site). I was denied access to all medics and some nurses – only those within peri-operative directorates were permitted (n=350). In brief, the original aim was impossible due to ethical constraints on both sites. As over a year had already been wasted awaiting ethical committee decisions (and a similar time span may be needed if pursuing further regions - perhaps with similar results), I limited the scope of the thesis so that data to be collected within the PhD timeframe (full effects of this decision will be explored later in the overall integration chapter). Hence a comparison between two different cohorts on two different sites (i.e. Eastern Nurses and Medics in the Midlands) will be made, and the scoping survey will be re-run using the new nursing sample. Nurses at the Eastern site were used to having posters in their work areas regarding research invitation, thus this method will be used in the main study. The

same generic questionnaire was drawn up but with demographics *added.* This was completed by the sample and proved to be very similar to both Midland samples (EBP nurses and medics) in that 'clinical application of physiological knowledge within the respiratory system' was top of their 'most wanted' list. When talking independently to appropriate education stakeholders, they too offered the same 'most wanted' list with no prior knowledge of what nurses/students had said. (This may indicate that the EBP group is actually more representative of nursing/healthcare populations than first imagined).

When surveys perform well Kelley (2003, p262) believes that the following advantages should be evidenced: i) empirical data gathered; ii) More generalisability to a population; and iii) large amounts of data can be produced a in a short time.

Empirical data gathered; The data collected was based on real-world observations. In contrast to common survey expectations, a large amount of 'rich' and theoretically relevant data was gained due to anticipating pre-survey that data collected was likely to lack details on the investigated topic. Thus on the grounds that the data was i) authentic and 'rich'; and iii) contextually relevant to the learners' competency, it can be said that empirical data was gained.

More generalisability to a population; Before generalisability can be discussed, survey differences require inspection. Using communication methods that participants were most familiar with provided an adequate response but this raised obvious 'comparability' issues between surveys (i.e. How much of the findings are related to the communication methods themselves?). All communication methods yielded very similar response rates indicating that the response is *unlikely* to be linked specifically to the method. This also adds weight to the notion that a reasonably high level of research 'saturation' existed in the target population (resulting in survey 'fatigue') before I came along. When using a new population (Eastern Nurses) in the main research, this showed a similar trend regarding both response rates and survey fatigue. When looking at other HEI (for similar surveys) response rates are comparable. This highlights that the usual 'state' for these populations may actually be 'chronically over-saturated'. It is therefore possible that the response achieved is actually close to the optimal response and is 'normal' for these populations. The ethics committees' restrictive permissions give further credence to this notion. Overall, the population findings may actually be more generalisable to the respective whole populations they represent than first anticipated. (Although relatively small numbers were gained n=141, their answers appear to be representative of their respective populations when tested later with new samples. This indicates may that learning preferences are similar enough in different cohorts regarding LO learning to allow them to be transferred easily from one sample to another). On the grounds that i) findings may be more generalisable than first anticipated, ii) learning preferences are similar enough to transfer easily between samples, it can be said that 'more generalisability to a population' was evidenced.

Large amounts of data can be produced a in a short time; This was certainly true. Once surveys were underway it took a total of 3 weeks to collect and analyse the data. In summary, all Kelley's (2003) survey advantages were evidenced.

Conclusions

The results provided a sufficiently good 'geography of the landscape' to provide a useful baseline of student learning desires and preferences. It is too early to tell yet but the indications are that the

surveys are potentially more representative of the whole population than first anticipated. Hence this will be discussed later in further depth. Response rate to the scoping surveys provided a good basis on which to build relevant LOs.

3.3. PART 2: Usability Studies/Grounded Theory

3.3.1. Usability Study

Creation of the LOs

Sims (2006) advocates that each media element should be assessed as to the extent it enhances learner engagement with the content which implies deliberate design choices. Thus I created the 7 LOs (based on Pathophysiology, Clinical examination skills, Normal Structure/function, and Infective/inflammatory/ metabolic processes of the Respiratory systems) to encourage user interaction. Kennedy (2004) believes that promoting cognitive activity and engagement enables the learner to gain better understanding of the relationship between content and context. Sims (2006, p4) supports this emphasising the importance for LO users to be "actively engaged with the environment... regardless of the developmental stage of the learner or their field of study...assuming that critical thinking in a flexible, complex... environment is the primary target". This engagement can only be achieved "if participants have the opportunity to test assumptions (hypothesize), construct solutions (manipulate), adjust variables (experiment) and/or introduce content (modify) within that environment" (Sims, 2006, p4). This freedom to 'test and engage' was therefore built into the LOs.

Usability Test Creation

Rationale

Having gained information about what medics and nurses wanted and created the LOs, a pilot study to expose any technical faults was needed. Although focus groups may have been a quicker way to report on tested LOs, the potential for strong characters to dominate group dynamics was not desirable. I wanted to harness individual thoughts for the LOs in order for them to be ultimately relevant to all. Regarding on-line learning Krug (2000, p141) states: "Focus groups are *not* good for learning about whether (it) works and how to improve it... or whether people can actually use it", hence usability studies were considered. There were several reasons why these were deemed 'suitable': i) they may promote feelings that technology is aiding the learning process i.e. not getting in the way; ii) learning is appropriate for the required context (Sims 2006, p4 supports this, adding that learners should be included in the design process and decisions); iii) usability studies could aid "a comprehensive understanding of how people learn and the way learners can best be engaged in online environments is needed" (Sims 2006, p6); iv) usability studies may help develop effective new models: "Where conventional instructional design models and process fail to effectively address (how people learn), we need to consider new models that integrate the pedagogies of online, learner-centred environments" (Sims 2006 p6); and v) I wanted to ensure that technological errors/problems

did not confuse results by becoming extraneous variables when testing approaches during main research.

Background to Usability Test Development

The following 4 perspectives helped shape my thinking behind usability study structure:

i) Some of Krug's reflections (Chapters 9-11, 2006) were considered as I wanted to know whether the LOs 'worked' and were 'user-appropriate'. I also wanted to allow unique areas (requiring testing) to come forth naturally. It was hoped that potentially inappropriate decisions regarding the testing approach would thus be avoided. As no research has yet been performed regarding LO testing in this context the 'best way' is unknown;

ii) Thomason (2004) believes that usability testing helps to reproduce the experience of the average user (important when using LOs at home on-line) and correct problems and outline useful aspects: "Do visitors enjoy using the LO and understand its purpose? (If so, they'll stay longer. If not, there's no compelling reason to return); Can users recover from errors? (Frustrated users are not likely to return – ever") (p4-5);

iii) Garrett (2003) believes that everything the user experiences should be the result of a conscious decision. This perspective again underlines the need for conscious design choices and was therefore adopted. When pursuing the wider on-line context Garrett also states: "By thinking about the user experience, breaking down into its component elements, and looking at it from several difference perspectives you can ensure that you know the ramifications of your decisions" (p19). This will be considered during usability analysis to unearth possible decisions ramifications;

iv) Garrett (2003, p23-5) outlines five 'planes' that provide "a conceptual framework for talking about user experience problems and the tools we use to solve them". These are: i) surface plane; ii) skeleton plane; iii) structure plane; iv) scope plane; and v) strategy plane (Appendix 11). Each plane is dependent on the preceding one - this created concerns. It is possible that correcting one problem may necessitate changes on several planes thereby forcing 'total LO revision' due to 'knock-on effects'. As this is an unsubstantiated concern Garrett's planes will tested.

Type of Testing

Krug (2006, p140-3) makes several valid points concerning the creation and conduct of usability studies and recommends 'Lost our lease' usability testing (Appendix 12) in order to be comprehensive, timely and appropriate; to make testers aware of how different people think; "Testing one user early in the project is better than testing 50 near the end" (p142); to show that "Recruiting representative users is overrated" (p142); "The point of testing is ...to inform your Judgment" (p142); and that testing should be iterative. Krug believes: "You'll always get more revealing results if you can find a way to observe users doing tasks that they have a hand in choosing" (2006, p152). For this reason, testers will be told that the 7 LO topics have been chosen by student doctors/nurses and that they have free choice regarding what they say, which LO/how many they test, and the order they do them in. It was hoped

that not only should this provide more revealing results and greater 'testing ownership', but would also mirror what that students might do naturally when using LOs at home.

Usability Questionnaires

The usability questionnaire should capture the participant's overall perception of the LOs usability (Kaufman 2006). Krug says that if 2 testers make the same mistake "don't let the third one make the same mistake – it will needlessly embarrass them" (Krug 2006, p152). However, Kaufman (2006, p5) takes a more relaxed view: "Let the participant make mistakes. This will reveal aspects of the interface that may need improvement". I decided to allow testers to make mistakes and observe how they attempted to solve them as this would most naturally mirror what would happen to students learning asynchronously from home – i.e. no help would be at hand, just themselves and the computer. If users became irretrievably 'stuck' I would intervene with help/reassurance.

Number of Testers

Nielsen and Landauer (2000) have shown that testing 5 users tends to uncover 85% of the problems. They state: "There is a serious case of diminishing returns for testing additional users" (p148). Krug thinks the ideal number of users is 3 or 4 (due to the first 3 users being likely to encounter all of the most significant problems). Conversely, Kaufman (2006) recommends testing between 5 and 20 users. However, Krug's ultimate recommendation is to 'test twice with fewer users' (Figure 9):

Figure 9: Krug's '2 Round' Usability Testing (Taken from Krug 2000, p14)

ONE TEST: 8 USERS, PROBLEMS FOUND: 5	8 users may find more problems in a single test but the worst problems will usually keep them from getting far enough to encounter others
TWO TESTS: 3 USERS, PROBLEMS FOUND: 9	3 users may not find as many problems in a single test, but in the second test (after first set problems are fixed) they'll find problems they couldn't see in the first

His rationale is as follows: i) it increases clarity; ii) when the intended audience is split between clearly defined groups with divergent interests/needs, users from each group should be tested at least once; iii) "If... your site requires specific domain knowledge then you need to recruit people with that domain knowledge for at least one round of tests... (but) don't do it for every round if it will reduce the number of tests you do" (p148). In my case, specific knowledge (e.g. evaluation and pedagogy) is required for specific groups (e.g. nurses and medics) which would indicate that a larger first group would be required. There is much to commend Krug's approaches, however there are also contradictions regarding required numbers. Krug's '2-round' approach will be used but a final decision will be made 'post-test' as to whether Kaufman's or Krug's recommendations will be followed for the main research.

Tester 'Type'

Krug believes that *anyone* can be tested – "The best-kept secret of usability testing is the extent to which *it doesn't much matter who you test*" (p152). It was intended that testers would be drawn from medicine/nursing but to test Krug's theory they will also be drawn from non-representative groups –

e.g. science education, information technology and 'laypersons' (i.e. those who have no scientific knowledge at all). To test the '2-round' approach, the first round will be representative testers (medicine) and the second round will be non-representative (laypersons).

Correcting Problems Approach

Two authors suggest ways of correcting problems:

i) Krug outlines three typical problems: i) Users are unclear on the concept; ii) words they're looking for aren't there; iii) too much going on. Krug advises: i) Ignoring cases where users momentarily 'wander' but recover almost immediately without help; ii) not adding anything when concepts are not understood – "the right solution is to take something away" (p148); iii) Taking 'new feature' requests lightly: "*they* wouldn't be likely to switch; they're just telling you what they like" (p148); iv) fixing obvious problems immediately (i.e. 'triaging'); v) implementing changes that are highly visible but require least effort. Krug reminds researchers that minor changes can have major impact;
ii) Garrett's (2003, p23) '5 planes' notion makes reference to user experience problems and tools used to solve them. Initially, this seems a straightforward way to correct problems by simply 'tracing back' to the offending articles. However, there appears to be several flaws: i) if an early wrong decision is made (or a problem realised retrospectively) on one of the foundation/initial planes it is likely that the whole LO would be affected due to its hierarchical nature. This may lead to a situation where; ii) the research schedule may be affected unduly (due to time required to fix flaws); iii) if many LO changes requiring a high degree of personal judgment this could potentially become 'what I think will work' rather than 'what actually does work'.

Since there are perceived limitations concerning Garrett's method and the types of required corrections are as yet unknown, Krug's problem correction method will be adopted. This appears more time-effective, affords a greater degree of flexibility (important during main research development), and appears closer to the aim of being fully user-centred.

Recruitment

Both Krug (2006, p149) and Kaufman's recommendations for *recruitment* were followed i.e. the incentive of helping to create useful LOs was offered), testing was conducted in rooms very close to the testers place of study (no travel), the participation invitation was simple, and a neutral testing location was used. "Contrary to popular belief, you don't need recording equipment or data-logging software...to run casual tests with a small group of users on an iterative design all you need is a system to test, a desk, two chairs and a participant" (Kauffman 2006, p3). Thus, this was exactly how the tests were conducted.

Usability Observation/Interview Technique

Ryan & Bernard (2003) show that textual data with verbatim text *without* rich narratives is sufficient to discover theory (i.e. brief descriptions of 1-2 paragraphs - providing that usability questions have been given sufficient thought as to provoke appropriate, 'rigorous' answers). Cohen et al (2000) also

emphasise that 'users are integral to the data's meaning' therefore verbatim text was chosen to produce theory to be tested. All comments will be recorded and where usability questions are not answered naturally by these comments, they will form open questions to be asked during a semi-structured interview (to cover all LO usability aspects). Kvale (1996, p180) recommends that interviews should be transcribed according to the most appropriate method for the research. In order not to intrude on observations, real-time written recordings of verbatim comments will be performed. Observation notes will include learning environment, route taken through learning, student misconceptions and overall LO use.

Creation of Usability Paperwork

Agimo (2004) suggest using the usability checklist from their 'User Profiling and Testing Toolkit'. As no specific LO usability questionnaire existed (and as this was the most comprehensive checklist available) I used this for the usability testing (Appendix 13). Components deemed important for all e-learning (according to Krug 2006, and TLT 2010) were added with a few modifications. The following questions were added: 'Is the LO 'fit-for-purpose'? Does the technology encroach on learning? What aspects are most effective/enjoyable? What does 'good' learning include? What would you like to see changed?' The usability checklist (Appendix 14), participant information and consent sheets (Appendix 15) were drafted. As users verbatim comments were needed, I also designed a fieldwork sheet (Appendix 16). Some recommend using a test script (Krug 2000, Kaufman 2006) but do not say why apart from making test delivery easier. I wanted a standardised delivery to ensure that extraneous variables were kept to a minimum. Krug's (2000) script was similar to what I wanted to convey so this was used after appropriate modifications (Figure 10).

Figure 10:- Krug's (2000) test script

"Hi, my name's Davina and I'll be going through this session with you today. You probably already know what we're doing - we're testing learning objects to see what it's like to use them. Even though questions may be asked after the test, we're testing the usability of the learning object, not you. You can't do anything wrong or make mistakes, and we would like to hear exactly what you think. You can't hurt our feelings, the more critical you can be (either positive or negative) the better. We need to know honestly what you think so that packages can be developed. You will be able to choose which learning object/s you most want to use (you may choose more than one - this will aid the researcher if you do so, but do not need to do more than one if you do not wish to). You may complete as many or few as you like. As we go along, I'm going to ask you to think out loud and tell me what's going through your mind as you go through the pages. If you have questions, please ask. If I cannot answer them right away (due to possibly biasing the research) you will be told this and they will be answered as soon as the testing ends.
You must click onto the next page using the blue arrow; failure to do this will mean that you end up in a 'blind alley'. With your permission, I will record what happens and what you say verbatim. This will be only used to help figure out how to improve the learning, and it will only be seen by others if necessary. After hearing this and reading the information sheet you are happy to take part, please sign the consent form. This indicates that you agree to participate and that we have your permission to record any comments you may make. Do you have any questions?"

Ethical Approval

Ethical approval was sought and granted for medical students but was restricted to advertising on one poster in an isolated corridor. After one year of trying, not a single participant had been recruited. This

was relayed to the committee who immediately extended approval giving me virtually 'Carte Blanche' access thereafter. Thus VLE was used to invite participation.

Usability Testing and Results

When conducting usability testing it rapidly became apparent that observations with interviews would be an excellent way to gather the intended new data for GT theory generation. To test this notion, I designed and used a data collation sheet to put answers into a simple format for coding and collation (Appendix 17).

Number/type of tester

Earlier it was noted that if specific knowledge is required for specific groups then i) a larger usability group will be required; ii) one round of testing (5-8 'representative' users) should be followed by package modification; iii) a second round of testing should recruit users until no new problems are found. 5 medical students were recruited as first round testers - only one technical problem was found but they desired several changes spanning all of Garrett's 5 'planes'. The 5 second round testers were laypersons. By the second user it was obviously that no new problems were being found, however I continued with 3 more users 'to make sure'. No new problems were discovered.
These findings showed that laypersons, although able to spot obvious problems pertaining to the first 2 planes (i.e. basic text, images or button problems), could not/did not comment on structure, scope or strategy planes. Krug's theory regarding non-representative testers was true in as far as laypersons spotted the same technical fault as the medical students, but false if graded or more in depth LO feedback is desired (as in this thesis). Laypersons had served a purpose in ensuring that the LOs were 'layperson-friendly' but will not be used for the main research testing later on due to the desire for in-depth LO feedback. Secondly, although '2 round testing' supports an iterative process (one that has been used with LO's for a while – e.g. Cook's work) the second round of testing did not highlight any further problems hence 'one-round testing' will be conducted in the main research. To ensure that 'problem saturation' is reached in the main research, Kaufman's recommendations will be followed. This means that the number of testers recruited will need to continue until saturation is reached. The data collection period may have to be left 'open' as a result.

Observation/interviews

Observations and interview technique were deemed 'good' - in-depth answers were obtained highlighting the potential for this method to be used for grounded theory.

Analysis methods/Correcting problems approach

One technical fault during testing was identified (an image resting over text after animation). Before fixing this, both Krug and Garrett's methods were compared. When using Krug's method it was not as straightforward as anticipated. The sum total of what it reveals was as follows: i) no users were unclear on the concept; ii) a couple of users felt that a glossary would be a good idea (perhaps indicating that essential explanations were not there or users did not have a basic understanding of

terminology); and iii) In a couple of instances users felt that a page had too much writing. This does not reveal much and leaves out important preferences – e.g. navigation. Although problems were triaged according to Krug's method, this took a fair amount of time and the necessary remedial actions were not immediately obvious. In short, Krug's approach *appeared* more straightforward, but in this instance proved unwieldy. I therefore tried Garrett's method. Far from causing unmanageable and time-consuming problems, fixing each level clearly did *not* mean having to completely re-write the dependent levels *in reality* as I had wrongly assumed. Even though Garrett says that levels are dependent on each other the feared hierarchical knock-on effects were undoubtedly misjudged as the framework easily identified how many levels were affected (in this case most of them as larger pictures, a glossary, and more back buttons 'test functions' were desired). Surprisingly, Garrett's method proved to be a useful tool when considering LO modifications and will therefore be used in the main research study if required.

After the technical fault was corrected changes that users wanted were added. I decided that the LOs should have 'blind alleys' built into them to observe what users (when they are not following instructions) do when they get 'stuck' to ascertain what percentage of users do not read instructions and to what extent LOs should be 'foolproof'.

Overall Conclusions

Proposed overall usability testing, observation and interviewing techniques proved to be fit-for-purpose therefore no modifications apart from 'minor tweaking' was necessary. Without exception all participants said that the LOs were enjoyable and highly relevant to their needs.

4. PHASE 1 MAIN RESEARCH - PROJECT 1A

The aim of this chapter is to explore the underlying pedagogical and evaluation reasons as to why certain approaches are effective for LO learning. It has two parts: Part A - Multi-site usability studies (to provide new source data for grounded theory), and Part B – Grounded theory (to develop the emerging theory to be tested, verified and further developed in Phase 2).

Two sites/populations will be used - i) The Midlands-based medics population performed well in preliminary work so this population was retained for Phase 1A; ii) As stated in the preliminary section, no demographical data was forthcoming for the EBP population thus no direct comparisons between year group breakdown, locations nor disciplines could be made (thus no generalisability estimate was possible). This Midlands nursing site was therefore abandoned and a further site pursued (Eastern). As Dumas and Redish (1999) recommend a varied 'spread' of testers, participants will be recruited from these backgrounds as before.

The planned process will be delineated first followed by both sites being discussed together. The findings will then be compared and broken down where appropriate in order to identify similarities, differences and to outline grounded theory parameters. Both populations are subjected to usability studies and Grounded theory, and comparisons made.

4.1. Multi-site Usability Studies

4.1.1. Rationale

As previously discussed, usability studies were tested in the preliminary work and found to be a good and appropriate method to gather the source material, and for hypothesis testing for further development. The rationale remained the same. As stated in the preliminary work, although the 'normal' kind of usability studies can be given a set number of participants to use them and valid data gained, it is anticipated that data collection will remain open until the grounded theory aspect of the approach is fulfilled to ensure that saturation has been reached.

4.1.2. Creation

As the paperwork performed well in preliminary work the same paperwork was used here (with the exception of questionnaire amendments (i.e. questions added – Do LOs do what you expect them to? How does technology enhance LO learning? What aspects of the learning are most effective or enjoyable? What does good learning include? What would you like to see changed?).

As before, the usability study consisted of i) observation of the participants use and choice of the LOs during field observation (verbatim comments and observation of use were completed on the data collection sheet – Appendices 16, 18 & 20); ii) post package questionnaire and iii) semi-structured interview (using answers given in the questionnaire to gain a greater understanding of answers given). It was felt that fully structured interviews (despite being able to statistically to prove/disprove

hypotheses) would limit the findings to my own understanding of the topic. This was therefore not appropriate. Totally unstructured interviews would risk usability questions not being entirely covered and consequently risk inadvertent measuring of extraneous 'technical' variables. It could be argued that this necessary covering of questions (i.e. semi-structured interviewing) may have weakened the 'pure' grounded theory method but this stance is defended on several levels:

i) Glaser's (1998 & 2003) edict was followed religiously: "In a face-to-face interviews... the researcher is advised to speak as little as possible and when necessary, further questions are asked and are limited to what has already been said e.g. "Tell me more about...".
ii) If an ethnographic study had have been undertaken fully unstructured interviews would have been conducted. As previously mentioned this was not desired, but enough depth to produce hypotheses was. Hypotheses were easily produced from the semi-structured interviews and did not seem to weaken the outcome in any way (see later in this section);
iii) Grounded theory hypotheses needed to be later tested alongside systematic review findings during main research testing. This was made possible because the semi-structured interviews prevented a potential total mismatch between the two. Consequently, both 'what' is effective and 'why' were found (see Main research: Phase 2, Chapter 7);
iv) Holloway & Wheeler (2010, p102) state that "observation is not only complementary to interviewing but is also a form of within-method triangulation". Thus my observation would flag up any mismatches between behaviour and answers given in the interview.

With regard to optimum number of participants for usability studies, Krug (2006) states that there is no point testing more than 9 people once or 3 to 4 people asked twice as no new findings will occur. Nielsen and Landauer (2000) state that 5 users tend to uncover 85% of a site's usability problems. Although it was not a site being tested, 'web' versus 'material to be used on the web' have clear similarities. Virzi (1992) found that 80% were detected using 5 participants and 90% using 10. Thus, as 80%> was seen to be i) a high percentage; ii) adequate percentage to be effective (as shown in preliminary work); iii) it would be difficult to get participants in twice due to time constraints; and iv) 5 users per discipline would be practically feasible, it was decided that at least 5 users per discipline per LO (coming once) should be tested giving a population recruitment target of 35 per site. It was accepted that saturation may occur before, at or after this number. As such, a decision will be made regarding the final number needed at the time of collection, i.e. usability testing will continue until both saturation has occurred and no new usability problems present themselves.

The intended target audience was a combination of participants drawn from nursing and medicine. Nursing and medicine were targeted because these were the disciplines under study. Science teachers were included for 'related discipline' comparisons, IT to ensure that there were no major faults in the design that novices would not spot, and laypersons to test Krug's theory that anyone should be able to use it. If similar principles were present across all group types then generic principles may be present. It was intended that participants from both sites would be recruited, and that the LOs would be tested on randomised and blinded participants using 7 specially designed LOs (developed during preliminary work according to the stated needs and desires of students from both disciplines) to see

which types of approach were most effective. Inclusion criteria consisted of adults aged 18-58 years from the stated disciplines of all levels of learning.

4.1.3. Data Collection

Due to a lack of theoretical saturation (and lack of required number of IT participants) during the research 'window' time, recruitment was left open slightly longer than anticipated in order to get the required amount (hence some other disciplines recruited more than the required 5 participants). A total of 57 participants were tested on at least one LO each (and who were recruited over the two geographical sites). Field notes were taken from participant observations regarding 7 different Learning Objects (LOs) which were collated with participants' verbatim comments.

4.1.4. Results

General findings were that all participants from both sites i) felt that navigation through LOs were good, ii) liked the overall look of the LOs, ii) felt the content was clear & well-organised, and iii) images loaded well. 12.5% wanted extra links. Differences found in the Laypersons group were that they were i) the most critical group regarding the LOs, *ii)* noticed pictures first, *iii)* Men 'hopped around' the LOs, whereas women took a more linear route. In the Science teacher group: i) Titles were noticed first, and ii) there were less comments on the LO pedagogy. There were no specific differences in the IT group but the nursing group noticed images first (25% suggested inclusion of white paper links, etc).
Specific findings were that differences between disciplines were due to learning preferences *not* pedagogy/evaluation issues and therefore deemed negligible. Negative points raised were minor and easily remedied (i.e. size of text in some places).

4.1.5. Discussion and Analysis of Results

The five questions added to the usability questionnaire from preliminary work testing appeared to be useful at this stage (and later proved to be very enlightening - see Phase 2 and Overall integration chapters). Observation proved to be useful as a 'within method' triangulation tool (as described by Holloway & Wheeler 2010, p102) as direct comparisons could be made. Field notes were used to either confirm answers or question users further with regard to their the perceived underlying meanings that they attributed to various parts of the LOs and several mismatches were found on several occasions thus helping to 'clean' the data.
Nielsen & Landauer's (2000) '5 users per discipline' figure was adequate to find all the usability problems but insufficient to achieve saturation from a grounded theory perspective; on average an extra half a person per category per site was needed i.e. 55 recruits in total Two extra participants (n=57) were recruited to ensure saturation had been reached (see Grounded Theory section).
Differences between disciplines were due to learning preferences – this was evidenced in the way the participants expressed their wishes - this point will be discussed later (See overall integration chapter regarding learning preferences and the extent to which it is important to cater for them). Negative points raised were minor and easily remedied (i.e. size of some text, clearer navigation in some places,

better quality images, etc). These initial findings suggested that LO learning structure should feel 'intuitive' to the learner. This supports Krug's 'Don't make me think' approach to web learning.

4.1.6. Conclusions

These results showed exciting promise as the main findings were similar across different disciplines, locations, and in most cases, gender. Reasons why certain approaches are effective for LOs were not immediately obvious at this point, suggesting that underlying factors may be 'complex' or 'deeply buried'. However the questionnaire and interview techniques had elicited good quality answers on which to build grounded theory. The LOs tested appeared to be of sufficient quality to test evaluation methods providing that consideration for specific relevance to specific disciplines was incorporated.

4.2. GROUNDED THEORY (GT)

4.2.1. Rationale

The aim was to take all the usability findings (i.e. studies, observations and interviews) as source material and generate hypotheses/theory. Several GT methods were considered (i.e. Glaser, Straus, and Corbin). Glaser's viewpoint was closest to my own - Glaser believes that interviews are always necessary in order to truly validate the data from participants - doing a literature review before data collection 'contaminates' data. For these reasons I asked each participant on completion (of observation/usability test/interview) whether the verbatim comments/field notes written were a true reflection of what they had said/done. I also wanted to conduct the systematic review literature after conducting the GT and wanted a method that was capable of both identifying and testing theory. However, Glaser & Strauss (2009, p103) state: "The grounded theory process without constant comparison can only suggest theory which needs to be tested sometimes qualitatively but more often quantitatively". The creation of 'a theory needing to be tested' was desired. The Constant comparative method therefore had the ability to be used to i) develop grounded theory (using usability source material); ii) to develop the theory further (using the systematic review findings where possible later on); iii) Implement emergent theories in practice (i.e. in the main research study); and iv) Evaluate emergent theories and test them (in the main research study and ensuing discussion). These qualities were very much in keeping with my desires for this thesis so Glaser and Strauss' (1967 & 2009) GT approach using the constant comparative method was chosen.

Stern (1980) makes a case for grounded theory in situations where little is known about a topic or where a new approach capable of creating and developing emergent theory is required. As both of these aspects applied to the intended thesis research GT was considered as a potentially ideal method. "The grounded theory approach is a general methodology of analysis linked with data collection that uses a systematically applied set of methods to generate an inductive theory about a substantive area" (Glaser, 1998, p. 16). It generates concepts and relationships that "explain, account for and interpret the variation in behaviour in [the] substantive area under study" (p19).
One underlying desire for this thesis was that it should systematically gather empirical data – GT fitted with this desire. 'Classic' GT can be used with any type or mixture of data (Glaser, 2003, p83). As this

was appropriate for the inductive/deductive approach decided upon for this thesis (and because Glaser and Strauss' (1967 & 2009) approach using the constant comparative method worked so well in preliminary work), this approach was deemed 'fit for purpose'. N.B. Although GT has its roots in symbolic interactionism my thesis will *not* take this approach. The method has been chosen for its power to produce workable theories from source data which can then be tested but it will, however, consider the importance of the context in which people function (as in symbolic interactionism). Greater emphasis will be placed on whether approaches found credibly affect, compliment, or interfere with 'effective' learning. It will be interesting to see whether the approaches tested (i.e. pedagogies and evaluation) ultimately improve learning or just help to analyse individual actions and perceptions on the type of approach.

Research questions

The primary aim was to discover the core concept/s for LO research, pedagogy and evaluation. Hence GT research questions will be: 'How do effective approaches inform theory?' and 'How do efffective approaches inform practice?'

4.2.2. Planning the Grounded Theory

In order for the reader to make a judgment regarding the credibility and the appropriateness of the methodological rationale for this work, a full description of the method will be given.

Data Collection and Open Coding

All source data will be taken from the usability studies, observation field notes, and interviews and written on the usability data collection sheet. It is hoped that the data collected would be 'theoretically sensitive' in that it would give accurate meaning to and categorisation of the data when forming theories. Holloway and Wheeler (1996, p107) state: "...to be credible, the theory must have explanatory power... in a good project, linkages between categories and data are 'tight'. Therefore, I will strive to ensure that all potential theories have good linkage.

Developing a substantive code requires coding the data incident by incident (i.e. words/phrases). These can then be developed into concepts and the properties of these. Each incident is compared with each new incident and with emerging concepts and properties. Parahoo (2009, p6) echoes the sentiment of most GT texts: "Constant comparison is more than comparing scripts. It... means comparing emerging theory with similar phenomena". In other words, it includes thinking laterally about other things to shed light on an issue, and including things that challenge the emerging theory. Holloway and Wheeler (p106, 1996) recommend this method be used throughout the coding process. Open coding with a fairly high level of detail will therefore be used to conceptualise the data and form ideas. This will be taken through level I and II, and level III (axial) coding as described by Hutchinson (1986) and/or Ryan & Bernard (2003) (Appendix 18), Open coding will continue until the core category or categories are found and i) no new information on a category is found in spite of the attempt to

collect more data from a variety of sources; ii) the category has been described with all its properties, variations and processes; and iii) links between categories are firmly established.

Number of Data Units Needed, Selective Coding, Theoretical Sampling, Saturation and Memos

Kuzel (1992) advocates 6-8 data units (individuals/studies/sites) for homogenous groups and 12-20 for heterogeneous groups. It was not known what type of data would be produced as no GT work has ever been completed on this specific area. As such, the number of participants will depend upon how homo/heterogeneous the groups are. The core category (i.e. the emergent theory) is the one that links all others and is the core variable that will be added to the main research for testing. In order to increase population validity and ensure researcher bias and interpretation are kept to minimum, only minimal selective coding will be done (i.e. the linking all categories to core category/ies). Any findings that change the core category/ies (i.e. biases) will be listed as conditions and a search will be made to see if they occur in other groups (both those under study and groups to be tested later in this study). Any necessary modifications will be made, thereby developing the theory.

The constant comparative method has 4 stages –

i) Comparing incidents applicable to each category during coding - each time the same type of incident crops up it is compared to what has gone before (this 'comparable note-taking' is termed 'memo-writing');
ii) Integrating categories and their properties – ideas formed during this stage can be used to predict similar ones in a wider sphere;
iii) Delimiting the theory – fewer major modifications are needed as the process progresses. The researcher is quickly able to see whether items are a 'good category fit' or not;
iv) Writing the theory – as this suggests points 1 and 3 (with memos) form the content behind the categories (Glasner & Strauss 2009, p105).

Once this is complete, multiple data sources (anecdotal experiences and systematic review data) will be sought to develop the theory further (Glaser & Strauss 2009, p67 and 105).

Glaser & Strauss (1967) advocate theoretical sampling: "The theoretical ideas control the collection of data therefore researchers have to justify the inclusion of particular sampling units" (p77). Initially I felt that a high level of theoretical sampling may 'constrain' the data and wanted to see instead how the data developed. Parahoo would support this stance up to a point: "The decision (to use theoretical sampling) ultimately depends on the emerging hypotheses and on how important it is to produce theoretical completeness of groups or individuals" (p6), it concerns "those whose contribution can shed more light on refute or confirm emerging theoretical ideas" (Parahoo 2009, p6). The preliminary work was useful in that it showed that the method worked well for the chosen aim of the research (i.e. it was an appropriate method for developing emerging theory), however, it did not intimate what level of theoretical sampling may be necessary for this part of the research. Glasner & Straus (2009, p109) state that a risk of not doing theoretical sampling means that data can become unwieldy and hard to link therefore theoretical sampling and analysis should be performed simultaneously.

I wanted the data collected to define the need for theoretical sampling thus a modification of maximum variation sampling was chosen initially (Polit and Hungler 1999, p298). Nursing and medicine were

targeted because these were the disciplines under study. Science teachers were included for 'related discipline' comparisons, IT to ensure that there were no major faults in the design that novices would not spot, and laypersons to test Krug's theory that anyone should be able to use the learning. It was anticipated that there would be no further need for theoretical sampling in Phase 2 as further samples were desired to test the possible existence of generic principles further (see Phase 2 introduction). Thus data will be gathered as in preliminary work concerning not just pedagogy and evaluation but also regarding the processes around these concepts. Parahoo (2009, p6) warns: "The temptation to develop an all encompassing theory, especially in the confines of a doctoral study must be guarded against". There will *not* be an attempt to extrapolate an all-encompassing theory – but if one is readily evident it may be included and/or tested if time allows within the confines of this doctoral study. Furthermore, if 'unwieldiness' becomes a problem a greater level of theoretical sampling will need to be considered/adopted.

As stated, participant recruitment and data collection will continue until theoretical saturation has been reached. Glaser & Strauss (2009, p61) defines this in the following way: "One reaches theoretical saturation by joint collection and analysis of data. When one category is saturated, nothing remains but to go onto new groups for data on other categories and attempt to saturate these... also". Holloway and Wheeler (1996) assert that it does not always mean that saturation has been reached purely by using techniques like i) a certain word/phrase frequently mentioned, and/or ii) identifying the same ideas that arise repeatedly; but believe that Morse (1995) has a better way of evaluating whether saturation has been achieved or not. Morse states (1995, p149): "saturate data are rich, full and complete" – i.e. it is the quality rather than the quantity of data that is important. Glaser & Strauss (2009, p61) add "When saturation occurs, the analyst will usually find that some gap in his theory, especially in the major categories is almost, if not completely, filled". Coding will therefore continue until saturation point that is defined by no further properties being identified or when new incidents just provide more constituent parts of existing properties already coded. This point occurs at different points for each study and therefore cannot be predicted ahead of time (i.e. pre-study). Memos will be used where necessary to keep track of the emerging theory (Glaser, 1998, p177).

Glaser & Straus (2009, p62) believe that "It is important to contrast theoretical sampling based on the saturation of categories with statistical (random) sampling", but "Statistical tests of an association between variables are not necessary when the discovered associations between indices are used for suggesting hypotheses" (Glaser & Straus 2009, p200). As one aim of my thesis is to suggest hypotheses, these will be tested later and therefore statistical testing is not needed at this point.

Theoretical Coding, Sorting and Write-Up

Theoretical coding will be done by identifying the theoretical code(s) that explain how concepts relate to each other, and whether any related concepts could be possible 'cause and effects' . If this is found to be the case, hypotheses will be made. All relevant data (including literature) will be used to compare with gathered data during the grounded theory process, and during the coding stages – i.e. 'Evidence' will be collected from comparative groups to see if the grounded theory is similar to already existing evidence (Holloway and Wheeler 1996, p106). The final stage of Glaser & Strauss (1967) GT will be performed as follows to create a reliable set of theories/hypotheses: i) Comparing this with projects with large numbers and the same hypotheses; ii) Confirming instances and their conditions

(both within the study and compared to other studies); iii) Disproving instances and their conditions (both within the study and compared to other studies); Making central propositions, variables, and dimensions: and iv) Identifying situations that push variables to their limit i.e. do original effects hold true?

Theory Generation

According to Glaser and Straus (1967 & 2009), two types of theory are produced - i) substantive theory - one that applies to one context, with all topics related to the one under study; and ii) Formal theory – one that applies to multiple settings or describes the settings context (thereby placing it within a wider scale), and is compared to *all* topics - not necessarily related to the one under study. Glaser and Straus (2009, p237), believe that substantive and formal theories should have the following 4 properties: i) the theory should closely fit the substantive area in which it will be used; ii) must be understood by laymen; iii) must be sufficiently general to be applicable in a multitude of diverse daily activities; iv) it must allow the user partial control over changes that inevitably occur during daily use to ensure that it is worth using. Both formal theory and substantive theory are middle-range - they are not just hypotheses and not yet grand over-arching all-inclusive theories (p32). I hope to develop a substantive theory on which pedagogical and evaluation approaches can be tested further. Although not an official aim of this thesis, it is hoped that formal theory may be produced if time allows. If so, this too should be tested further.

4.2.3. Working Through the GT Stages: Results, Analysis and Discussion

The grounded theory data is presented here as a "theoretical discussion using conceptual categories and their properties" (Glaser and Strauss 2009, p31) rather than a well-codified set of propositions. This method puts an emphasis on process not just product. Throughout this section, it is hoped that the rigorous adherence to the method strived for is self-evident, and that the collection methods and data obtained were accurate and appropriate.

Data Collection: Open Coding, Constant Comparative Method and Sample

A full list of codes and categories are displayed in Appendix 19. Data was simplified into positive, negative and neutral comments (see Appendix 20). Initial open coding was performed as intended by labelling the items chronologically as they came up when going through the participant's answers on the questionnaires, interviews, and field note observations. These were entered on the devised paperwork. Each incident was compared with each new incident and with emerging concepts and properties. Incidents and memos were written on index cards.
Recruitment of participants and coding continued until no new codes/categories appeared, categories had major consequences and reasons for those consequences described (i.e. properties, variations and processes), and links were firmly established. Differences in categories were checked (minimised) in order to maximise the properties of each category. When data appeared complete, theoretical saturation was reached. (To check this, disproving instances were sought for each core category using

further new recruits from all types - education, health, science, IT, laypersons. No further instances were noted).

Glaser and Strauss' (1967 & 2009) GT approach using the constant comparative method worked well in this format. The paperwork and index cards proved to be a clear way of seeing the codes and determining which categories did not fit initially with any others. Thus this method will be used for the main research testing and the hypotheses produced tested further. The selection of the sample was seen to be appropriate as all Kuzel's (1992) 'good' sampling characteristics were evidenced.

Number of Data Units Needed, Selective Coding, Theoretical Sampling, Saturation and Memos

Data units - As the type of group (i.e. homogenous or heterogeneous) was unknown an anticipated initial aim was for 20 data units per group. In practice only 6 were needed, after which no new categories emerged. This indicated that groups were largely homogenous (so there is a high possibility that this number of data units were sufficient).

Selective coding - Codes were then grouped according to subject and whether the statement was positive, neutral or negative so that trends could be seen immediately (Kaufman 2006 p3).

Theoretical sampling and saturation - As categories/codes were a tight fit and the recruited sample seemed to cover all main codes purposeful sampling was not required. Theoretical sampling was done at the time of analysing "in order to discover categories and their properties and to suggest the inter-relationships into a theory" (Kaufman 2006, p4). Collection of data was not controlled but was allowed to 'happen'. The only time that this deviated was when a lack of IT participants was noted. Recruitment time was extended in response to this until the required numbers had been recruited. As no attempt to purposely hunt out IT participants was made, this meant that further participants from other areas were also recruited. This was justified as I wanted as little manipulation of the sample as possible. The fact that this data did not become unwieldy also justifies this approach. Concepts were fairly easy to link and data collected appeared to be 'theoretically sensitive' in that it appeared to give accurate meaning to data, categorisation of the data was straightforward, and instances were confirmed by further data.

Memoing - Kaufman (2006, p6) recommends entering all observations/questionnaire answers into an excel spreadsheet regardless of the type of observations made. Kaufman then recommends analysis by grouping similar observations (described by a short sentence and defining the problem and its impact on the user experience). If many groups are present "these can be organized by severity of the usability issue, translated into key findings, and trends extracted" (Kaufman 2006, p6). This simple method had much to commend it – i.e. defining the problem, and predicting trends in particular. It may also predict which methods have most impact. When trying out this method in practice however, I found it difficult to collate fields – the limited space in excel cells proved insufficient to record/capture the changing nature of concept development. Each concept was therefore left on index cards and new ideas/changing relationships were thus added and understood more easily. Hence, index cards will be used for the main research testing (Phase 2). Memos were sorted into piles according to concepts (Glaser, 1998, p187) and written up (Glaser, 2005b). They were integrated into the categories and

proved useful in keeping track of the emerging theory. Few modifications to categories were required during the 'delimiting theory' stage as most new incidents fitted well within existing categories.

Predominantly good descriptions of the major categories were found so the coding was deemed to be successful. Indicators on which these categories were based were drawn directly from the verbatim comments of the participants. The GT was a good fit, was relevant and easily modifiable. Categories and their properties complemented what was happening in practice with regard to the participants, the stakeholders and from my perspective. It was therefore deemed of good construction. The GT explained major variations in learning behaviour put here and therefore has achieved relevance (Glaser, 1992, p15).

Findings: General

General trends were easily evaluated using the positive-negative method and included the following:

- 70% of participants preferred information presented as images/animations whilst the remaining 30% preferred written/text elements (they liked summarised information, simple terminology/ definitions);
- 75% of participants liked 2 or more elements per screen to help focus learning and prevent boredom
- 50% of participants preferred the inclusion of test elements, the remaining 50% appeared to favour any method that aided the practical application of learning;
- Minor changes wanted in elements were equally distributed between visual (i.e. larger pictures) and written elements (i.e. less/more text, more labelling);
- 50% of participants felt navigation through the learning material on-screen was excellent. 42% felt that navigation was mostly good. 8% of participants felt navigation was inadequate but negative comments were minor (i.e. add overview, extra buttons according to preference);
- Comments about content of diagrams were all positive although a small percentage of participants (3%) wanted slightly larger images.

Findings: Underlying Factors

Strauss (1987) describes the main characteristics of core theories as: i) the central element of the research; ii) related to other categories; iii) explains variation in behaviour; iv) frequent reoccurrence in the data developing a pattern; v) easy/intuitive connection with other categories. These were evidenced in the data.

As coding progressed it became obvious that there were eleven core codes with regard to underlying pedagogical and evaluation factors: information overload, time, monotony, interest, choice, confidence, attention, control, application of learning, motivation and participant learning preferences.

Information overload – When the right speed and level of text/information was achieved the learners' confidence increased and levels of anxiety, confusion and the feeling of being rushed decreased.

Time – When well-timed information and layout was achieved enjoyment and the feeling that the learning was in line with their own learning preferences increased. Actual time taken to do the package decreased.

Monotony – This category was indirectly obvious through interest and attention categories.

Interest - Increased interest was achieved by increasing/varying stimulus, and having a good mix, level and clarity of information. This resulted in learners having their learning expectations fulfilled.

Choice – When navigation was clear, this increased learner perception of learning 'choices'.

Confidence – When speed of text and font was good, this increased learner confidence. If pedagogy style changed this produced decreased confidence in the learner.

Attention – When information and images were of high quality this increased attention by 'varying' the stimulus.

Control – This category was indirectly obvious through navigation (ISD) and layout (pedagogy).

Application of learning – When audio features were good this increased learner engagement due to the change in stimulus. When engagement increased learners found it easier to apply knowledge and their knowledge desire was satisfied.

Motivation – When text and information is good it 'hits the right level' for the students, and increases motivation to learn.

Participant learning preferences – Low user control over navigation (timing and layout of learning) increases the learners' feelings that the learning is not in line with their learning preferences.

Findings: Main Factors

Pedagogy – Several elements of 'good' pedagogy were found. Effective pedagogy contains

- The same pedagogy throughout – changing this creates a confidence 'crisis' in learners;
- Using appropriate learning 'channels' enables the learner to evaluate the learning 'worth';
- Mixed elements on each page have a wider learner appeal as learners believe that this encompasses more learning styles;
- Good learning objectives enable information to be deemed as 'good' by enabling learners to judge their learning progression. If these appear too fast learners feel rushed and forget that they are in control of the speed of learning;
- Lack of 'revisit' options decrease the feeling of learner control over navigation and increase the feeling that the learning is not in line with learner preferences;
- Good audio increases application of knowledge linkage to practice, raises engagement through varies stimulus and satisfies increased knowledge desire.

Evaluation – Findings indicate that evaluation can be made easier through clear layout and learning objectives. This gives the learner the ability to self-evaluate. The above underlying and main factors will be taken forward to Phase 2 and compared to Phase 2 findings in order to verify them using

different populations, disciplines and locations. There was no obvious single over-riding substantive code. This was not seen to be a problem as I wanted to leave categories deliberately 'wide' to stay as close to learners' use and perceptions as possible. Data collected was 'theoretically sensitive' and gave accurate meaning to and categorisation of the data. Categories/answers were checked with participants who verified that these were exact reflections of their experience. The theory was beginning to have explanatory power. Data was therefore highly valid and credible (Holloway and Wheeler, 1996, p107). Tight linkage between categories was seen but findings will be regarded as provisional (until they have finished being subjected to the constant comparison method and used in Phase 2 testing).

Subgroup Analyses

Similarities and differences between the 57 participants from nursing, medicine, science, IT backgrounds and 'laypersons' recruited were compared. Differences were not large. Laypersons were the most critical group regarding the LOs and they noticed pictures first. Men generally chose a non-linear route, whereas women took a linear route through the LOs. Science teachers noticed titles first and gave less comments on LO pedagogy. Both groups were largely similar in their LO use.

Theoretical Coding, Sorting and Write-up

A full list of codes and categories are displayed in Appendix 19 (total = 95 codes). Data simplified into positive, negative and neutral comments can be seen in Appendix 20. Data was broken down into levels of practice, disciplines, location and gender.

Following grounded theory generation it is clear that the questions (added from preliminary work testing onto the usability questionnaire) proved to be a crucial part with respect to the foundations of the grounded theory hypotheses (i.e. 'Is the learning in the learning objects appropriate/fit for purpose? Does the technology get in the way of learning? What aspects of the learning are most effective or enjoyable? What does good learning include?' and 'What would you like to see changed?'). These questions elicited well-reasoned answers and made tacit responses explicit.

Theory Generation

Having discovered general trends using the positive-negative method, and having discovered underlying pedagogical and evaluation factors using the constant comparative method, Glaser & Strauss's (2009, p196) method for generating theory was used. They state: "One comparative strategy for generating theory from findings is to compare clusters of relationships *within* the context of emerging theory". This was done by comparing associations within and between single questionnaire items relating to the same category and to the core index (Table 2 Pedagogy, Table 3 Evaluation):

Table 2: Comparing Associations Within and Between Single Questionnaire Items (Pedagogy)

Concept=pedagogy	Minimum = poor pedagogy	Maximum = good pedagogy
Consistency	Changing pedagogy creates 'crisis' of confidence in learners	Consistent pedagogy increases learning confidence
Learning channels	Bad mix of learning channels increases feeling of overload	Good mix of learning channels the learner can evaluate the worth of the learning
Mixed elements on each page	Learners believe that all learning styles are not catered for	Has a wide learning appeal as it is seen to encompass all learning styles
Revisit buttons	Lack of revisit buttons increase feeling that learner is in control of navigation and therefore learning	Right level of revisit buttons increases feeling that the learning is in line with personal preferences
Learning objectives	When learning objectives appear too fast learners feel rushed and forget that they are in control of the speed of learning	Good learning objectives enable information to be deemed as 'good' by enabling learners to judge their learning progression
Audio	Decreases engagement through lack of varied stimulus	Increases application of knowledge linkage to practice, and satisfies increased knowledge desire

Table 3: Comparing Associations Within and Between Single Questionnaire Items (Evaluation)

Concept = evaluation	Minimum = poor evaluation	Maximum = good evaluation
Layout	Evaluation is difficult when layout is not clear	When layout is clear evaluation is easier
Learning objectives	Learners feel unable to judge their progression	When learning objectives are clear this gives learners the ability to self-evaluate

Another comparative strategy for generating theory from findings is to compare clusters of relationships *between* different consistency indices:

Table 4: Comparing Clusters of Relationships Between Different Consistency Indices

Concept	Maximum=appropriate level	Minimum=not appropriate level
Information	Good speed/level of text/information increases confidence	Bad speed/level of text/information increases anxiety, confusion and feeling of being rushed
Time	Well-timed information and layout increases enjoyment and decreases time needed to do the package	Ill-timed information and layout increases feeling that the learning is not in line with personal learning preferences
Monotony	Increases learners perception of having their learning expectations fulfilled	Poor mix, level and clarity of information decreases stimulus to learn
Interest	Increased learning choice when navigation is clear	Increased perception of lack of learning choices
Choice	Good layout/navigation increases choice due to being well set out/clear	Bad navigation produces feelings of lack of learning choice (i.e. not in line with learner preferences)
Confidence	Increased confidence when speed of text and font is good	Crisis of confidence when pedagogy is not consistent
Attention	increased attention when information and images are of high quality	Decreased attention via lack of stimulus
Control	Good timing and layout of learning increases perception of control over learning	Bad timing and layout of learning Increases perception that user has low control over navigation, and that the learning is not in line with personal learning preferences
Application of learning	increased learner engagement when audio features are good	Increased difficulty in applying knowledge and decreased desire to have knowledge needs satisfied
Motivation	and increases motivation to learn When text and information is good	Decreased motivation when learning does not 'hit the right level'
Participant learning preferences	Good user control over timing and layout of learning increases perception of varied learning choices	Bad user control over timing and layout increases perception of limited learning choices

All theoretically relevant relationships among the concepts were extricated for each core index (pedagogy and evaluation) and their underlying factors.

Hypotheses

How concepts related to one another and possible cause and effect relationships were considered. Linked cause and effects formed the hypotheses (a full list is available on request). These will be tested in the main research testing phase 2 in conjunction with top performing pedagogies and evaluation approaches gained from the systematic review).

Hypotheses produced by the GT included:

- Mixed LO elements results in increased knowledge, interest (and achieves a good 'element-interaction' balance);
- Layout that is not in line with participant learning preferences results in section links not being detected;
- When information is of good quality learners are able to use learning objectives to judge the level of learning attained resulting in them believing that they have progressed;
- Labels giving the right amount of information result in clarity and simplicity that aids navigation and does not overload the participant.

Plotting The Theory

All codes, categories and comments were put onto pieces of paper and laid out onto a huge piece of paper. They were laid out in order of causative items with arrows (i.e. if appropriate level was seen to cause motivation an arrow was placed running from appropriate level to motivation). This continued until all codes and categories had been incorporated.

Core Category Comparison with Current Literature to Further Develop the Theory

Existing Evidence and How These Findings Link In

In line with Glaser and Strauss' method, comparative literature/knowledge/experience was collected during the coding stages to see if the GT was similar to already existing evidence (Holloway and Wheeler, 1996 p106). It is fair to say that there is a paucity of comparative literature.
General findings included '97% of participants liked 2+ elements per screen to help focus learning and prevent boredom'. This is consistent with a dual-process model of working memory. Learners construct links between words and corresponding images when they are held simultaneously in working memory (Mayer et al 1999). This is also supported by modality and contiguity principles. Moreno and Mayer (1999) did two experiments; in one learners viewed on-screen text with animation, in the other information was narrated with animation. Learning was measured by retention and transfer: Students learned better when image and text were geographically close (spatial-contiguity effect), and when information was narrated rather than when visual text given (modality effect).
This may hint at why 'good audio' was important as a major pedagogical factor in the findings. Good audio was seen to 'increase application of knowledge linkage to practice, raise engagement through varying stimuli and satisfying increased knowledge desire'. The differences between text and auditory

pedagogy may be explained by Britton et al (1998) in relation to meta-cognition. Learning from text is determined by making connections concerning the following aspects: metacognition (the ability to sense that a mental representation needs extra connections to become coherent); inference-making ability (this spreads memory activation within working memory to make necessary connections); and domain knowledge (provides prior knowledge for inference-making) – i.e. it is a complex process. Mayer & Moreno (1998) showed that text and images use the same learning channels, whereas auditory and visual learning channels are separate. Therefore the importance of using different learning channels (and not overloading the same channels) may be under-estimated. Mayer et al (2001) shows that 'channel overload' can happen when too much text and images are given at once. This was evidenced in my findings on one occasion when the learner described their feelings at being tested (deliberately) with a large, solid block of text on-screen.

Jurden (1995) linked working memory with speed factors. In both 'Information overload' and 'Time' categories 'correct speed of information' was important. The decrease in anxiety/confusion and time taken to use the LOs *and* the increase of enjoyment found in my research could be directly linked to the fact that the information has had opportunity to be retained in the working memory.

The links found between 'Monotony' and varied stimuli are well-established and therefore will not be discussed in detail here. This finding indicates that the importance of varying learning stimulus in LO learning is no different to other forms of learning. Similarly, although very few papers comment specifically on LO learning objectives, it is well documented that appropriate and measurable learning objectives result in effective learning.

No specific examples of 'Good labelling equals lack of overload' were found in the literature, however Steve Krug's (2006) publication 'Don't make me think' highlights this statement indirectly. It suggests that *intuitive* e-learning designs provide the most effective pedagogies.

Missing Literature

The links evidenced between 'Timing', 'learning styles/preferences' and 'control' do not seem to be adequately covered by existing evidence. It will therefore be interesting to see how this area progresses with further testing planned for the main research Phase 2.

Implications for Quantitative Research (1B & Phase 2)

Implications are that a good linkage between GT and the systematic reviews will be crucial. As no bias was desired in the writing of the systematic review protocol this was completed before GT analysis was performed. As such, there is a risk that links between systematic review and GT findings may be difficult. However, it is assumed that if generic principles are at work in defining which pedagogical or evaluation approaches are effective for LO learning then links will be relatively obvious providing that the systematic reviews are conducted well. It is also important for reasons of parity that the GT methodology used in Phase 2 is the same as that used here. In this way direct comparisons can be made. The same Midlands-based population of medical students (but not the same participants) will therefore be used to verify/refute the theory when tested blind. To develop the theory further, Glaser and Strauss advocate testing other groups/populations. This will be done in main research testing

phase 2 with a new nursing population in order to test the theory in terms of whether it holds true when discipline and location is changed. At this stage it is difficult to see what implication for LO provision/providers, technical designers, learners, and tutors may be. This will therefore be addressed in Chapter 1C and incorporated into the main research testing Phase 2.

It could be argued that this GT did not achieve 'full' saturation in that Glaser & Strauss (1967)'s five prepositions were not fully met i.e. comparison with projects with large numbers and the same hypotheses. However, this was not possible as none presently exist. Rather than being a fault of this work, this strongly emphasises its originality. This is further supported by i) experts in the field have been very interested in the findings of this thesis, thereby suggesting that the findings are indeed 'cutting edge'; ii) many well-documented published theories relate indirectly to matters concerned with this research (e.g. multi-modal learning within e-learning increases knowledge retention, interaction and participant interest (which appears to particularly support the 'Mixed elements' finding); and iii) Confirming and disproving instances and their conditions within the study, and central propositions, variables, and dimensions were found. This was important. Any hypotheses created may later need this capability if the formation of generic principles were possible. It was also important as principles needed to be contextually relevant but allow a level of standardisation in order to be in line with present educational benchmarks.

Systematic Review Data (that has a bearing on the core theory hypotheses)

No systematic reviews currently exist on effective LO pedagogy or evaluation. As previously stated, I wanted to compare these findings with the systematic review findings *when completed.* As it was not known at this point the extent to which systematic review findings would dispute, complement or even triangulate these GT findings it was important generate theory and hypotheses so that bias could be kept to a minimum before being taken forward and tested in the main research Phase 2. It was also important that that should be done before any systematic reviews meta/statistical analysis to prevent bias/type I error. The GT produced here is not considered to be 'complete' until the systematic reviews (Stage 1B) have been conducted and the findings compared with this GT, i.e. comparing 'what' with 'why' certain LO pedagogies/evaluation are effective in practice. The extent to which they integrate will be discussed in chapter 1C. References to Glaser and Strauss' (1967 & 2009) two main types of theory will not be discussed at this point as the full weight of what type of theory this produces must be put into context with the systematic review findings and any research developed from this later in this thesis. This will therefore only be discussed in chapters following 1B.

As previously stated, I desired that GT should not be biased by a literature review (necessary for the systematic review protocol). Therefore to prevent bias and be consistent with the GT method (Glaser 1998, p68), GT source data for the grounded theory should be decided upon and conducted before writing the systematic review protocol (i.e. performed sequentially). However a mixture of 'political' problems and highly restrictive ethical approval in one of the sites unexpectedly rendered the planned time scale needed for separate and sequential gathering of GT data and systematic review evidence totally unworkable. As a result, a slight overlap between the grounded theory and the systematic review data collection was forced upon me. Although this is not ideal, it should be noted that all 'source data' research for the grounded theory had been collected and coding/analysis completed before finishing the protocol and conducting systematic reviews. As I was able to keep methodological

integrity, this unanticipated problem is not believed to have affected the research, but it will however be discussed in later in the overall discussion section.

Hypothesis testing

The next GT stage incorporates 'hypothesis testing' - i.e. i) Projects with large numbers and same hypothesis; ii) Confirming instances and their conditions; iii) Disproving instances and their conditions; iv) Central propositions, variables, & dimensions; and v) Situations that push variables to their limit (i.e. whether original effects hold true). There are presently no other projects in existence, let alone any with large numbers and same hypothesis. Hypothesis testing will be done in main research testing phase 2 Chapter 7 at which point multiple data sources (including anecdotal 'lived' experiences, and systematic review data once completed) will develop the theory further.

Comparing GT with other groups

Glaser & Strauss (1990, pp47-54) suggest comparing dissimilar groups (i.e. one substantive group to another, or comparing several subgroups) to develop emerging formal theory. This has already been done using different types of participants i.e. subgroups. To develop the theory further this will be first verified/refuted by using different participants in the same population in medicine, and then further theory developed by verifying/refuting the GT by using different participants in a different population – nursing (i.e. comparing dissimilar substantive groups, one to another). If this is appropriate and links core categories are similar, this may legitimately extend the concept and scope of LOs from the smaller area of 'medicine' to a larger one i.e. 'health'.
In summary, comparison of groups provides i) control over both conceptual and population generality; and ii) simultaneous maximisation or minimisation of both the differences and similarities of data (these are both vital for developing emerging theory, Glaser & Strauss – 1990 - p55). More work on minimising differences between comparison groups needs to be done to i) bring to the fore any data not previously/currently found; and ii) to establish a definite set of conditions under which a category exists – which in turn ii) establishes a probability for theoretical prediction. N.B. Care will be taken to qualify and not just broaden the research/categories.

Answering Research Questions

As all work has not yet been completed all comparisons to develop the theory are not yet possible. Elaboration analysis, if appropriate, will be discussed after completion of 1B and Phase 2 testing. Work on Project 1A/Grounded theory provided sufficient evidence to comment on the inital GT research questions. The question 'How do effective approaches inform theory' was not easy to answer. Several interwoven factors play a part. An example from my data will be used to illustrate the answer – e.g. when learners do not follow instructions they lose their digital sense of direction. It was obvious from the observation of participants that navigation then becomes confusing and learners are more likely to give up if working alone. Added to this, when layout is not in line with participant learning preferences this results in section links not being detected. I tentatively wondered whether inserting specific page positions on each page may have circumvented the need for all learning styles to be catered for. Although, this appears to be true on face value, without additional research this cannot be commented

on further at this stage but will be commented on later in the thesis. What is clear, however, is that effective approaches must have clear instructions; a failsafe mechanism whereby learners can easily find their way again if they get lost due to not following instructions; and a constant reminder on the present page of where they are overall. This is also supported by the fairly large number of people asking for exactly these things during inital usability testing in preliminery work. It would be logical to assume that having these things in place leads to the perception that instructions are clear, navigation is easy and a sense of confidence concerning where they are in the overall package. However, perceptions of the learners were found to be 'greater' than this – i.e. learner perceptions not only encompassed all these things but also felt that the learning had been of 'good quality' and 'did not overload the learner'. This is an important finding as effective approaches appear to inform theory *by telling the researcher specifically what the 'value added' learner perceptions are*, i.e. those not *immediately discernable.* This may also explain why lack of various elements appear to cause the learner to believe that the LO learning is not in line with their personal learning preferences (e.g. timing of information and layout). It is noted that to create LOs that contain optimum levels of personal preferences for all learners at all times is akin to providing the impossible. It is possible that pedagogical design *may not have to* provide the impossible if timing of information and layout is 'good'. This concept will be further examined once 1B and Phase 2 have been completed. Other examples of this include:

i) When labels give 'appropriate' amounts of information this does not just result in clarity and simplicity, but also aids navigation and does not overload the participant.
ii) When 'mixed' elements (i.e. more than one element, e.g. words/text, or sound/text) are placed on the screen, this not only increases knowledge and interest during LO use, but also achieves a good 'element-interaction' balance for the learner;
iii) When learning objectives are used, the learner is not only able to judge the level of learning attained, but they believe that the information given is good quality and that they have progressed.

Thus, it is the mixture of factors interacting together *within* effective approaches which then render implicit value-added learner perceptions explicit, which then informs theory. The question 'How do effective approaches/frameworks inform practice?' is more straightforward: they appear to inform practice by simply showing what works and what doesn't. They also consolidate and give greater or lesser emphasis to different parts of the approaches therefore exposing theoretical gaps. This allows educators to estimate how appropriate particular approaches are when applied to the chosen context.

Conclusions

'Value-added' learner perceptions have provided several indications as to why pedagogical development may have been difficult in the past, and these will be further developed during the course of the thesis. The evidence found here appears to provide an excellent basis for further development. This chapter's aim was to find out what underlying relationships contribute to effective pedagogy and evaluation approaches for LO learning. It did this. This thesis has evidenced the development of Grounded Theory based on a collection of primary data, and augmented by secondary data from the literature (and later the systematic reviews). The result is theory, grounded in empirical data. At this stage it is too early to assert that the evidence is 'accurate beyond a doubt' (in the words of Glaser and Strauss 2009), however the issues discussed are undoubtedly relevant theoretical abstractions about 'what is going on'.

5. PHASE 1: MAIN RESEARCH - PROJECT 1B

Here the main aim is to discover 'what' works in practice regarding pedagogy and evaluation (as previously stated, this was not undertaken until usability findings were collected, and grounded theory was underway). The Systematic Review (SR) method will primarily act as an effective approach assessment/evaluation tool but it will also as a 'hypothesis generating tool' (as described by Thompson 1994 and Verstraete 2002).
A detailed rationale for using SRs is outlined, alongside a summary of the background literature, existing evidence gaps, development challenges and barriers to progress. As there were no appropriate or adequate 'off-the-shelf' protocols in existence I needed to create a General Protocol Methodology using questions/sections that were generic in nature. This will also be outlined.

Three SRs are required. One to discover the type/level of existing LO research, one to find effective pedagogies, and one to find effective evaluation approaches. In order to be highly relevant (both theoretically and in practice) each SR requires the creation of additional questions specific to each topic. These also will be outlined, termed 'unique protocols', and added to the generic protocol. Results from both unique and generic protocols will be detailed culminating in directions for Phase 2 research testing, implications of SR findings and an evaluation the protocol efficacy.

The SR definition used here is 'a review striving to comprehensively identify all literature on a topic providing an overview of *all* relevant studies'. It will contain explicit and reproducible objectives, materials, methods and meta-analysis if appropriate. All efforts will be made to ensure that it has replicable methodology, clear reporting, and transparent and appropriate synthesis (Mulrow 1994, p499). As this area is new, it is possible that 'research types not previously envisaged' may be discovered. For this reason (and because this protocol is not yet 'validated'), the first five papers will act as a pilot study to ensure the protocol is 'appropriate' and identifies any missing categories.

5.1. Rationale

The basic *methodological* rationale for the SR format (as the deductive method of choice) has already been outlined in the introduction section. However, as the SR format is not an obvious choice a deeper rationale (with reference to the literature) will now be given to support my reasoning.

Reason 1: Lack of Empirical Evidence

In 1998 Ehrenberger & Murray noted that most LO evidence was anecdotal. To date, only two published papers *claim* to be empirical: i) Inglis et al (1999) and ii) Weller et al's (2003). Inglis et al (1999) advocated 'best practice' as the optimum method for dealing with education's transitional nature, asserting that design is informed by learning research. Howard-Rose et al (2004) support this seeing research as a means to support the learning design as a whole. They believe that 'systematic literature appraisal' aids course *decision-making* (and therefore LO design) but do not explicitly specify how this can be achieved. SRs *could* have the ability to aid decision-making by extracting learning outcome

data from different research methods. Christiansen & Anderson (2004) believe that Weller et al's (2003) outline for course design "provides excellent empirical analysis" (p2). Design seems to be a key element for LO learning, however they admit that "many of these benefits...are not necessarily related to *an LO approach* to course design" (p2). Their paper has many such incongruities and confounding variables have not been fully considered. As this is one of the *better* examples empirical evidence in the literature appears woefully inadequate.

West et al (2002) advocate *reviews* for providing empirical evidence. Davis (2005, pp77-79) recommends examining the evidence and integrating research and practice. Davis explores 'other professions' research use to inform the major discipline under question – i.e. research-based nursing practice. Interestingly, this takes the same approach as the one I use in this thesis. I had decided that considering all disciplines (at least initially) rather than just medicine/nursing alone would place any later findings within a greater 'learning' context (and possibly highlight important *generic* LO principles). Thus the first reason for conducting SRs was to assemble all existing empirical evidence into a synthesized format having the secondary effect of informing LO course design and decision-making.

Reason 2: Lack of an LO Evidence Base

As empirical LO evidence was sparse, not surprisingly there was no LO evidence base. Initially, LO advantages were seen nationally to remedy many pitfalls encountered by web-based learning. However, LOs themselves encountered problems and lack of national standards hampered progress. Additionally, robust research evidence was needed to inform clinical decisions, ensure best practice and increase delivery reliability. Green (2005, p270) believes that: "Systematic reviews aim to inform and facilitate this process through research synthesis of multiple studies, enabling increased and efficient access to evidence". Additionally, there appears to be many parallels between present LO evidence-base needs (rigor development) and Psychology/Medical Education over the last 40 years: i) 1970's - Psychologists highlighted the systematic steps needed to minimise biases/random errors in research reviews, and in 1987 Mulrow highlighted the poor quality of research reviews. Recently similar comments have been levelled at educational reviews; ii) 1988 - Oxman & Guyatt published guidelines to aid review quality and establish the healthcare evidence base. This directly reflects LO learning's present need; iii) 1992 - Antman et al expedited 'evidence base formation' when stating that resuscitation treatments were in existence *long after* they had been proved harmful. Although consequences for education are not as *directly* harmful, SRs on LO learning could gather robust research to strengthen practice-theory bonds, and perhaps provide a more effective/satisfying learning experience. In 1993, the BMJ and the UK Cochrane Centre examined 'the science of reviewing' and the importance of SRs within medicine was fully realised. The resultant guidelines are now well established for quantitative research.

Several developments over the last eleven years are noted: i) 'The first link between the 'development of education' and 'clinical practice' was made by Van Der Vleuten et al by highlighting the need for the same level of academic scrutiny in clinical practice to be present in educational practice; ii) There is a consensus emerging about the need for systematic reviews covering selected topics in medical *education*' (BEME 2000); iii) Wolf (2000) alluded to the 'practice and promise of evidence-based education' and described lessons that could be learned from evidence-based medicine; iv) medical education's effect on education: "The success of evidence-based medicine has led to pressure to make

medical education more evidence based... (but) good randomised controlled trials in education are hard to find" (Greenhalgh et al 2003, p109); v) Davis (2005) states a main role for SRs/meta-analyses is to facilitate evidence-based practice changes. More and more links between evidence-based practice in health and education were being made.

In 2005 BEME acknowledged this need believing that a guidelines for optimal LO learning practice would enable educators to develop the LO evidence base further. Prosser (2005, p8) appears to support this view: "There is a growing debate in higher education about evidence-based or evidence-informed approach to... improving... learning". He concluded that i) education needs to learn from Medicine; ii) 'evidence-informed practice is both possible and urgently needed'; iii) there is a need to gather and evaluate both qualitative and quantitative evidence; v) a different way of thinking in education is required; and vi) concludes with asking whether SRs are the way forward.

This presents a reasoned case for using SRs for LOs, however education RCTs are often impossible due to its complex inherent nature (i.e. confounding variables). SRs of educational research can identify good ways of learning but this will not always be applicable to all contexts, situations, learners or teachers. It must be judged alongside delivery, students' learning experience, and learning outcomes. Given these caveats, SRs historical development within medical education may suggest that a 'suitability assessment' (followed by careful review planning) has the potential to establish a robust evidence base for LO learning. Thus, a second reason for robust SRs is the urgent need to establish the evidence base.

Reason 3: What the Literature Means Regarding Evaluation/Pedagogical Practice

Many educators encounter difficulties when attempting to meaningfully arrange LO content. There may be several interconnected reasons for this: i) they have limited time available; ii) content is diverse; iii) literature is disparate; iv) specific LO literature is minimal; and v) underlying LO literature is large. Consequently, appraisal of *what the literature means regarding evaluation/pedagogical practices* becomes extremely difficult and the relevance of e-learning literature is not immediately obvious.

LO Research - When conducting a scoping literature review in 2004 for an LO research SR, only 21 published papers were found showing that 62.5% of literature was 'grey' (i.e. unpublished/presented orally). Importantly, only 12 papers contained enough information to form any kind of robust LO basis. The SR premise was reconsidered and possible reasons for this lack of rigor sought. This showed an obvious lack of clarity in the way studies were designed/evaluated/reported. As the method *per se* seemed appropriate (and there was a large amount of grey literature) SRs were feasible providing that greater clarity regarding design/evaluation is obtained from the authors.

Evaluation - The plethora of disparate information makes it extremely difficult for educators to pull together valuable 'knowledge strands' to evaluate on-line innovations: "Many innovations involving knowledge media have not been *appropriately* evaluated. Many have not been evaluated *at all*" (Inglis et al 1999, p157). Despite a rapid growth in university web use, literature searches show that evaluations are surprisingly scarce and often demand 'non-traditional' approaches (Bain 1999, Hagan & Markham 2000, Lie & Cano 2001, Postema & Markham 2002, Sheard and Markham 2005). By 2006 many HEIs had adopted LOs but few had discovered archetypal uses (due to being at 'initial production'

stages). Despite five years passing little development has been seen. Several reasons are offered: i) Lack of resources/expertise caused a lack of evaluation (Alexander 1999). Although much improved now a lack of *robust* LO evaluation remains. This is hampered by ii) inherent on-line complexities: "Standard evaluation techniques are not suitable for this complex technology" (May 2001, p92), and "The literature did not reveal a consistent starting point on appropriate methodologies with which to carry out such evaluations" (Sheard and Markham, 2005 p257). Consistent starting points may only exist when an evidence base is established; iii) A scarcity of practical evaluation guidance has hindered the development of practical evaluation guidelines (Sheard and Markham 2005); and iv) "Often evaluations of educational innovations involving digital technology originate from the innovators themselves" (Inglis et al 1999, p157).This shows that SRs may aid evaluation development (by establishing common LO learning themes) *if* independent evaluation of knowledge media is made.

Pedagogy - There is evidence in the current literature that E-learning is increasingly becoming part of the 'core business' of educational institutions. Back in 1997 the Dearing report suggested that internet technology was crucial to a 'learning society' (NCIHE 1997). As e-learning is now credited with 'having the potential to improve learning quality' (Gilbert et al 2007) online LOs form part of this. Winters et al (2010) highlight another important issue: "Despite... policy drivers, many institutions are struggling to embed e-learning effectively and much remains to be learnt about how technology can best be used to enhance student learning" (Winters et al 2010, p71). They recognise that although many HEIs are committed, more information regarding students' experiences of 'e-pedagogy' is required. This suggests that pedagogical experiences are crucial if a full understanding of LO pedagogy is to be gained. Unsurprisingly a literature search failed to reveal a commonly accepted LO pedagogical approach which according to Rossi et al (1999) would provide an established methodology for evaluators. Thus, a third reason for LO SRs is the urgent need to understand the literature to facilitate systematic evaluation/pedagogical practices. SRs may aid this process by establishing rigorous dimensions. Innovations can then be appraised thereby developing academic rigor.

Reason 4: 'Pedagogy-Led' Versus 'Technology-Led' Learning

In 2004 the educational, research and academic climate found LOs suffering from a pedagogical 'identity crisis' (Calbraith 2009). In 2006 I sat on the CETL RLO steering committee and noted that educators were having trouble developing pedagogy. Today, little has changed. 'Coal-face' educators remain unsure as to how to build LO pedagogies, what they should look like, or how they should be used/tested. Historically this may be due to the literature, research studies, and/or educators and web-developers adopting a stance in one of two completely different camps i.e. concentrating on technical software development (with pedagogy as a secondary aim), or on pedagogy (hoping that the software can present the ideas as planned). Both approaches have potential problems. An overemphasis on technology-led learning (i.e. gaming, AI) can lead to difficulty in applying the most effective/appropriate pedagogy (to enable the learner to achieve the learning objectives). Conversely, concentrating on pedagogy may induce tutors to create learning problems that software is incapable of executing. Therefore "Comprehensive evaluation of web-based learning environments need to consider both the technical... (and) pedagogical aspects of the system" (May 2001, p95). This 'linkage' difficulty may be overcome by SRs through the discovery of educational outcomes themes/patterns when using

particular technologies which can then be developed into optimum uses. This integration may subsequently aid the demarcation of effective learning parameters. Thus, the fourth reason for LO SRs is the urgent need to intrinsically integrate both the ideals and practice of technology *and* education.

Reason 5: Customised Learning and Quality Testing

In 1991, Inglis et al noted that the demand for educational 'quality' had never been greater. Today, this is vital. Students expect relevant courses that are sensitive/specific to their needs. To this end, the Higher Education Academy published 'the student experience' (2006), inadvertently reinforcing this 'consumer-driven' trend. 2004-2006 saw prolific LO production, partly in response to the increased customised learning need. However, prolific customised LO production requires educators to have the technical, pedagogical, experiential, and operational knowledge to effectively design/produce materials. CARET, UCeL, and CETL aimed to help more educators to develop these skills. In 2003, Duncan (cited Littlejohn 2003) stated that learners wanted to customise what they buy/use 'to make highly relevant courses'. My experience as a HEI Senior Lecturer supports this view. Duncan also believed that customisation was possible due to more educators having the necessary LO expertise. Although this has grown they remain relatively few. Keeping abreast with web developments and their potential educational uses seems a large problem which is likely to increase in the near future as more open source software, more technical advances and greater accessibility to broadband becomes available. Littlejohn (2003, p150) noted: "As the number of resources for networked learning continues to grow steadily, it becomes increasingly pressing that knowledge of these resources should be made widely available and the reuse of those resources by other groups should be facilitated". Alongside this, Krauss and Ally (2005, p19) highlights the "need to assure educators that they are using resources that are highly rated and represent some added value to the learner". Educators must also be able to evaluate the theoretical, philosophical, conceptual, and practical rigor. Further advances have been made alongside RLOs: IEEE's national standards, CAREO/MERLOT's digital repositories, Conole et al's in-house RLO framework, and meta-tagging parameters. These factors, combined with a need to develop professionally, have led to 'traditional' education being reviewed (Burns & Glen 2000, DoH 1999, GMC 1993, Ramsden 1992, UKCC 1999). However, a full evaluation of how these factors influence educational practice within LO learning have not yet been undertaken. Thus the fifth reason for LO SRs is the need to test materials for quality purposes and to facilitate systematic evaluation practices (due to LO-related advances).

5.2 Background Literature: Evidence Gaps, Barriers and Challenges

As previously stated, 3 SRs were required: LO research (to uncover what had been done, how, and whether this elevated certain methods above others); LO evaluation (to find top-performing approaches); and LO pedagogy (to find top-performing approaches). Pertinent background literature will be explored before discussing reviews to expose existing/potential gaps, barriers and challenges.

LO Research Gaps

As stated, no empirical evaluation of *research-based* LO pedagogy or evaluation approaches existed. Most previous work was principally anecdotal and large knowledge gaps existed within the areas of general healthcare, information technology (IT) and LO development.

General Healthcare - Until 1998 it was generally accepted that most clinical practice was ritualised, that the quality was variable, and that practitioners often did not make adequate use of available research (Ford & Walsh 1994, Haines & Donald 1998, Newman et al 1998, ENB 1998). In conjunction with this dilemma, several educational ideals were becoming important healthcare curricula attributes - experiential learning, lifelong learning, and inter-professional learning (GMC 1993, NCIHE 1997, UKCC 1999). These ideals were driven by several factors: a) government (DoH 1999-2002); b) professional bodies (GMC 1993, UKCC 1999, UKCC 2001, GMC 2002); and c) strategists (NCIHE 1997). This, in part, led to educators exploring new learner-centred delivery techniques/applications resulting in the realisation that traditional methods did not always bridge the practice-theory gap (Waddell 1991, Davis et al 1995, Wood 1998, CRD 1999). Much work over the past decade has been focused on 'narrowing' this perceived gap with greater adherence to evidence-based care/education.
In short, evidence had been gathered to outline the need for new learner-centred delivery techniques and applications, and Medicine, Nursing and Education had partially 'narrowed' the practice-theory gap. However, today very little has been done to bridge what I call the emerging *'techno-professional'* gap i.e. the gap between LO technology and professional learning (obviously, one PhD cannot hope to fill such a gap, but may add knowledge to lessen this emerging 'chasm').

General IT evidence - In the early 1990's many institutions developed CAL packages as part of their students' learning experience. However, these courses tended to be context-bound resulting in limited use. Later, Web-based courses appeared but they followed the 'monolithic' model due to pragmatic reasons (cost, lack of expertise/technology standards: Duncan 2003/Littlejohn 2003). As time progressed it became obvious that i) full realisation of many educators being able to create LOs had not been evidenced, and ii) learning resources were being duplicated – even within the same department. Thus the concept and impetus behind RLOs was born. Initially these factors, combined with a need to develop professionally, seem to have led to 'traditional' professional education being reviewed (Burns & Glen 2000, DoH 1999, GMC 1993, Ramsden 1992, UKCC 1999). Hodgins (2002) and Christiansen and Anderson (2004) advocate 'object-orientated' designs in which digital learning content is designed for 'modular formats'. However, this concept has been viewed as a 'Lego block' metaphor for course construction and has been criticised by Wiley who favours "a molecular model in which only certain atoms (LOs) can be combined to create stable molecules (units and courses)" (1999a, webpage).
In short, evidence concerning the need for 'traditional' professional education to be reviewed in the light of technological advances had been gathered, with some headway towards optimal construction. However, very little development has been performed concerning the actual *value* of LOs or their pedagogy/evaluation.

LO development evidence - Many authors have supplied definitions, characteristics, potential use, size theories, taxonomies, and some means of evaluation (Wiley 1999a & b, Downes 2000, Longmire 2000,

Wiley 2000, 2000a, b, c, d, e & f, Wiley Recker & Gibbons 2000a and 2000b, Muzio Heins & Mundell 2001, Gibbons, Nelson & Richards 2002, Hodgins 2002, Martinez 2002, Orril 2002, Rogers 2002, Williams 2002, Littlejohn 2003, McGreal in press cited Christiansen & Anderson 2003, Naidu 2002, Olivier & Liber 2003, Thorpe Kubiak & Thorpe 2003, Weller Pegler & Mason 2003, Wiley 2003). This said there is no direct evidence concerning what 'LO learning' *should look like*. Despite this lack, on-line LO learning indirectly assumes many things. It assumes that the learner i) has a computer; ii) knows how to use it; iii) can get a server connection when needed, iv) can afford to pay for the internet/broadband link; v) has enough keyboard skills to engage with this medium; vi) is able to learn adequately using this medium; and vii) is able to get enough access to a computer (given family demand, etc.). In HEI/professional training it also indirectly assumes that the evaluation, pedagogical and adult learning theories behind LOs are robust, valid, reliable, tested, appropriate, and able to inspire 'deep' and 'lifelong' learning. As such, many unanswered questions arise: Do learners have adequate skills to cope successfully with computers? How is LO learning best performed/assessed? What are each individual organisations needs? What are the individuals learning/development needs? What are the resulting effects on patients? How do HEIs compare in their LO use? It is hoped that a clearer picture of these issues may be gained throughout this thesis (see Chapter 9).

In short, evidence had been gathered to describe/categorise LOs better, but there is none regarding what LO learning should look like. Given the underlying assumptions and resulting questions it now seems pressing that answers are found. This thesis will therefore aim to delineate 'LO learning'.

Potential Barriers/Challenges to LO Research Progress

No previous empirical evaluation of LO research has been done. Preceding passages have alluded to several potential progress challenges – the need to bridge the new 'techno-professional' gap, measure LOs worth and delineate what LO learning 'looks like'. The actual value of LOs and LO pedagogy/evaluation will be weighed both separately (content) and in how they interact with surrounding factors (context). Despite my intuitive feeling that LOs have intrinsic worth, this may be incorrect. Factors around LOs may not influence their effectiveness (constituting a potential barrier). Although findings of this nature would be disappointing, LOs are so under-researched that uncovering a lack of effect would still advance the knowledge field. Should this be the case, greater examination of LO research and its processes will be conducted. In short, many potential barriers and challenges are currently unknown. Those identified centre around filling literature knowledge gaps.

LO Pedagogies

Evidence Gaps

The increasing interest in LOs (particularly *RLOs*) led to many new initiatives, government funding and hot debates between pedagogical and technical stances. In practice, 'in-house' design and evaluation formats (with intuitive designs and high face validity) were used, generally without robust investigation into the foundations on which to optimally build. Confusing information/practice has complicated the place LO pedagogy holds - some educators have attempted to develop LO theory/practice only to find

that pedagogical dilemmas and barriers 'bar their way' (Felix 2005, Calbraith & Dennick 2009). As the thesis progresses greater understanding of these barriers will be sought. Notably, three educators/institutions have made some headway:

i) It is suggested that Deepwell (2002) used 'Linear' strategies (i.e. whilst building an evaluation framework, she developed pedagogy sequentially in one direction) which encountered development problems during further development. When asking her why the learning had been constructed thus, Deepwell explained 'it seemed to be the best way to do it at the time'. Wang's (2008, p6) statement reflects my own experience and may explain Deepwell's pedagogical barrier: "Pedagogical designers usually use their experiences and tacit knowledge of assessment outcomes to design pedagogies". Perhaps 'tacit product knowledge' can only go 'so far' when trying to develop effective LO pedagogy;

ii) Others have employed constructivist strategies. It is suggested that Boyle (2002) attempted a *deconstructive/reconstructi*vist strategy (pedagogies were developed by dismantling/rebuilding existing pedagogies with high 'face validity'). Alternatively, Nash (2005) and Koohang & Harman (2005) tried to build from the 'bottom-up'. Jones & Boyle (2007, p26) believe that "LO designs... based upon constructivist principles could be excellent sources of design patterns that would enable LO authors to more easily produce pedagogically sound learning resources. This clearly suggests that constructivism/design patterns are the way forward. However, constructivist methods also met with limited success due to pedagogical dilemmas blocking further development (Calbraith & Dennick 2009). When asking Boyle about this, Boyle explained that many of his LOs had been designed not just for their pedagogical value but to increase student attendance;

iii) The New London Group (2000) used a 'multi-literacies' pedagogy involving four components that were later built on by Mills (2006) using situated practice (building on learners real-life experiences); overt instruction (guiding learners use of explicit design meta-language); critical framing (encouraging learners to interpret social context and purpose of meaning); and transformed practice (learners design/transform existing meanings to new meanings). This pedagogy apparently worked until its context was changed causing difficulties for students "who were not Anglo-Australian or middle-class" (p70 – Mills 2006). This indicates that this pedagogy may contain effective components but greater examination of its use/structure is required.

When Boyle and Deepwell tried to develop LO pedagogy they met with insurmountable conceptual barriers preventing or compromising further pedagogical development. They could not adequately express why further development was not possible thus neither approach provides a full answer as to how effective pedagogies should be designed. However, SRs may aid the 'design decision' process. The New London Group (2000)/Mill's (2006) pedagogy appeared to go further before hitting pedagogical barriers - it was effective until it used as a linear hierarchy or in distinct stages. However, reasons for barriers were unexplained. I felt strongly that 'lack of explanations' did not amount to an 'impossibility of 'pushing pedagogical development further', and that many explanations were possible – e.g. Deepwell's 'best guess' approach may be comparable with any other method in that it may fortuitously 'hit' on some effective LO elements but not others. Equally, when using a de/reconstructivist approach perhaps dividing the aim (as Boyle did) was enough to introduce pedagogical development barriers, or perhaps the 'sum of the parts' add up to 'more/less than the whole'. If so, this may explain why a simple dismantling/rebuilding process becomes inadequate for pedagogical development.

Conversely, a totally new/innovative approach may be warranted i.e. developing pedagogy alongside (instead of seperately to) evaluation. I felt that all of these possibilities merited consideration via SRs. According to Wang (2008) ten instructional method types are commonly used in pedagogical design – i.e. Presentation, Demonstration, Discussion, Drill-and-practice, Tutorial, Cooperative learning, Gaming, Simulation, Discovery, and Problem solving. However, they admit that use in education "might bring up new instructional methods beyond this list" (p5). This remains to be seen and will be discussed later.

Traditionally IT has taken Instructional Design/Constructivist approaches to research. Conversely, education has usually taken more formative/narrative approaches. Since LOs need to be firmly rooted in both education and IT (and neither approach appears adequate alone), this implies that a new approach may be required. The relative immaturity of academic rigor within e-learning has compelled many educators and researchers to take a 'fresh look' at LO/E-learning approaches resulting in the realisation that a more systematic approach is urgently required (Neumeier 2005).

In 2006, after the SRs in this chapter were conducted, Rohse and Anderson (2006) brought new promise when they pointed out that the literature was becoming littered with "digital technologies offering a means for realising complex pedagogies that free formal education from some of the constraints of the past" (p90). However, this hope was short-lived. Later in 2006 they stated "These insights have influenced recent learning design theory and discourse...yet there is a sense that this potential to realise complex pedagogies is mostly unmet". There could be several other reasons for this: i) Educators are still having problems developing LO pedagogical theory/practice due to conceptual barriers (Calbraith & Dennick 2009); ii) teachers' practices have fallen short of their 'espoused goals' (Prosser and Trigwell,1997); iii) existing evidence regarding the previous point has not been published; iv) it may not have occurred to Educators that specific LO pedagogy exists; or v) more information is required on which to base pedagogical assessment/evaluations. (These ideas will be considered during the research and re-evaluated in Chapter 9).

Potential Barriers/Challenges Concerning Pedagogical Progress

Apparent 'clustered' enthusiasm and government backing enabled several bodies to be set up, produce, monitor and evaluate RLO quality (e.g. CARET, CETL). This led some to believe (Christiansen & Anderson 2004, Whalley 2006) that LOs (RLOs in particular) could revolutionise learning. Yet, despite pronounced government backing during 2004-6 few answers were forthcoming. Reasons for this need highlighting. Other progress barriers include the vast amount of confusing information/practice still complicating the indistinct picture regarding LO pedagogy, and the amount of intuitive 'in-house' formats being used in practice without robust investigation. This thesis hopes to uncover reasons why pedagogical 'blocks' exist (and therefore overcome them) and guide educators LO construction.

Several challenges are evident: i) methods to enable complex pedagogies to be realised need 'unearthing'; ii) a consideration of whether teachers' practices meet their espoused goals is required; iii) new pedagogies may be needed as 'conventional' instructional design models and processes "fail to effectively address online, learner-centered environments" (Sims 2006, p6) However, this may not be possible due to unknown factors; iv) Rohse and Anderson (2006) submit that teachers' practices are

complex, contextually dependent on teaching environment constraints and subject to individual variation and that learners are encouraged to be independent, autonomous and self-directed. This sets further challenges – a) being contextually relevant without being constrained by the teaching environment. Teachers may be resistant/unable to meet espoused goals or contextually relevant pedagogies may not be effective unless constrained by the teaching environment; b) being contextually relevant whilst also having a level of standardisation (relevant pedagogies may be incongruent); and c) having the capability to allow autonomous, self-directed study.

LO Evaluation

Evidence Gaps

As previously stated, CETL/government funding backed LO development and Brown (2007, p26) notes: "Ideally, in order to maximise return on investment, we need to... reliably measure the probable learning outcomes of specific activities before significant resources have been invested in their development". Wang (2008, p6) believes that "the use of assessment outcomes to improve pedagogical design is underreported... because traditional educational systems do not store and utilize much assessment outcomes". Brown supports this stance: "It follows... that we need to specify the intended learning outcomes in advance and use this specification as a benchmark for testing the design as it develops" (2007, p26). Hence *outcomes* rather than *process* have been evaluated: "The available models for evaluation including theory-driven evaluation tend to be associated with clear and definable outcomes and have limited availability within the development of process tools" (Sheard & Markham 2005, p355).

Lack of ways to *measure* LO learning may also have contributed to most evaluation research concentrating on LO outcomes. One such example is Scriven's formative-summative model (1980) where formative and summative methods are based on evaluation needs/outcomes rather than methodological systems. Scriven describes a flexible approach where formative materials become summative as a new phase of evaluation begins. This involves an iterative process (that accumulates evaluation data within a systematic framework) but it is not necessarily *readily adaptable* to *changing evaluation needs*. This is fine if particular outcomes are desired, but the ability to be flexible (as well as making LO assessment and evaluation easier) was considered a key issue for this thesis. As specific reasons for evaluation difficulties do not yet exist, evaluation *processes* as well as outcomes will be examined. Sheard & Markham (2005, p355) advocate: "A further consideration in the evaluation of any web-based learning environment is that the evaluation must encompass not only the educational process but also the process associated with the functional usability of the technology". Thus: "There is then a need to explore possible models of evaluation that allow flexibility and sensitivity to this complexity" (p355).

According to Howard-Rose and Harrigan (2003), several multimedia evaluation approaches are evident in the literature: "consumer-oriented, expertise-oriented, objectives-oriented and participant-oriented approaches discussed by Nesbit, Belfer and Vargo (2002), Worthen, Sanders and Fitzpatrick (1997)... and Williams (2002)" (webpage). Williams' (2002) participant-oriented approach describes four key components: context, input, process and product. This further supports the need to include both

process and product. Williams (2002) believes that this approach is ideal for evaluating LOs "because it honours differences among various stakeholders' definitions of an LO and appropriate criteria for assessing its value" (p177). Nesbit et al (2002) use a participation model for LO evaluation where different stakeholder groups are used. It is a two-cycle process where participants evaluate LOs asynchronously. Worthen, Sanders and Fitzpatrick (1997) distinguish between consumer, expertise, objectives, and participant-oriented evaluations. However, because a *systematic* sweep of the literature has not yet been undertaken the *most appropriate* LO evaluation/processes/products are unknown. There is a danger that LOs may be no different to any other multimedia regarding appropriate evaluation approach, thus the 'market' may be more 'saturated' than expected. In short, LO evaluation processes require assessment and the following questions require answers: Are existing LO evaluation processes readily responsive to changing evaluation needs? Do they need to be? Do they encompass both the educational and functional usability process? Do they explore models that allow flexibility/sensitivity to the above complexities? To what extent are outcomes necessary for evaluation models? Do models have limited availability when developing process tools?

Potential Barriers and Challenges Concerning Evaluative Progress

To gain further insight into the above questions I undertook a quick examination of what had already been done in the general e-learning literature with regard to evaluation (see Table 5).

Table 5: General E-Learning Evaluation Work

Evaluation Framework Title	Approach	Comment
SESL - Systematic Evaluation for Stakeholder Learning: Ramage 1997	As appropriate	No clear or precise evaluation guidance
U21G - Universitas 21 Global	Problem-centric learning	Not content-specific nor student-centred
Untitled - Britain & Liber 1999	Laurillard's Conversational Framework 1993	Limits evaluation to individual students/tutors
Cost-aware Evaluation - Ash 2000	Integrated evaluation	Only okay for quality assurance if embedded into curriculum
Multi-method evaluation - Anderson et al 2000	As appropriate	Not specific
Integrative Evaluation	Jones et al's 1996 Integrative evaluation	Acknowledges context but has general correlation problems
Untitled - Joyes 2000	As appropriate	Acknowledges context but allows extraneous variables to creep in
Multi-institutional intelligent tutoring tool evaluation - Kinshuk et al 2000	Byzantium CILE model	Intended to test in the real environment (using others work)
CIAO - Scanlon et al	Jones et al's Integrative evaluation	Both qualitative/quantitative methods used
Untitled - Williams 2000	Oliver & Conole 1998 - Illuminative model, Countenance model, action evaluation model	Results are biased/skewed towards context
Holistic Evaluation - Shaw & Corazzi 2000	Illuminative - Oliver & Conole 1998	Iterative development is shown
Interactive Framework - Deepwell 2002	Oliver & Conole 1998 - Illuminative model, Countenance model, action evaluation model	Unwieldy and confusing
CINEMA 2003	Not stated	Poor navigation between course /evaluation system
Untitled - Conole & Smith 2002	Utilisation-focused evaluation - Patton 1997	Raises the question - What new pedagogies are possible?

It was clear that three main points had not been addressed:

i) Evaluation frameworks were not systematic (e.g. they were either 'not student-centred' or 'student-centred to the exclusion of all else', they had problems relating to the context or limited the evaluation to certain individuals);
ii) Generic evaluation frameworks had not been addressed, and
iii) nor had the generation of transferable models.

This seemed to result in e-learning being under-researched and untested (Sharpe & Benfield 2005), and evaluation suffering from the 'Shangri-la syndrome' as described by Williams & Goldberg (2005), i.e. people talk/dream about on-line evaluation but don't know how to get there or if it really exists. It was clear that drawing upon the general e-learning evaluation literature was going to be difficult and may not help me to gain specific LO evaluation insight. Approaches that were systematic, generic and transferable were needed. It is suggested that an approach that *combines* both educational and functional usability processes may aid this. Therefore the first potential barrier may be that an approach that *combines* both may not exist or even be possible.

Secondly, although direct evaluation of tutor practices may illustrate whether LO evaluation approaches 'work' regardless of tutor, this had already been attempted and had failed during preliminary work. This concurs with other researchers experiences of direct educator interaction on this topic. Difficult collation of direct evaluation practices is therefore a potential problem. If direct collation in Phase 2 is impossible the SR may alleviate this need by collating evaluation practices indirectly.

Thirdly, difficult collation may influence important evaluation practices/ideas assessment. Some educators have addressed this by using non-traditional approaches (Hagan & Markham 2000, Postema & Markham 2002). Collation difficulties may prohibit any useful development/conclusions regarding effective evaluation approaches using traditional methods and new ones may need development. Fourthly, non-traditional method development may be impossible for reasons unknown as so little work has been performed to date. LOs are 'under-researched' but *LO evaluation data* appears critical to their current/future viability for several reasons: i) formative evaluation ('user-testing'/self-evaluation) during instructional implementation provides feedback to instructors/developers and can enhance students' learning experience quality; ii) instructors seeking to use/re-purpose others LOs need to assess LO usability; and iii) evaluation data offers evidence on LOs developed by different faculties.

Three Educationalists/Researchers attempted to find answers: i) Trigano & Pacurar (2004) used Gagne's Learning Theories, Instructional Design Theory, Elaboration Theory, Merrill's (1999) research, and Reigluth's (1999) research. Their method is similar to LAMS Activity Management; ii) Knight, Gasevic & Richards (2006) used an ontology-based framework to house pedagogic learning designs. It links conceptual models with tools but pedagogy is conceptually abstracted from content and context; and iii) Es & Koper 2006 developed a learning design specification based on the IMS model (Expert/document analysis and Learning Design Coding but not an evaluation model per se). However, these suffered from the same problems of being unsystematic, non-generic or non-transferable. In addition, their work highlights many pedagogical barriers – i.e. the topic area is too big, too many

perspectives and too many disciplines to take into account; Topic area is too time consuming; Too much is not yet known; Too much data to sift through; Difficult to extricate the multiple levels of evaluation required/appropriate methodologies/appropriate pedagogies. In summary, the potential barriers/challenges uncovered by these background literature overviews partially explains why the anticipated 'mass roll-out' has since been nationally unfeasible on the scale originally imagined. Much had been tried with little reward. Remaining researchers (e.g. UCEL) consequently focused on granularity (e.g. GLOs).

I believe that LOs educational value may be *grossly under-estimated* due to developmental barriers. LOs may have much to commend them when i) barriers are overcome and challenges met; ii) LO intrinsic worth has been evaluated; iii) effective pedagogy/evaluation approaches have been researched. SR methodology (together with 1A) may illuminate understanding with regard to these. The first challenge is obviously to overcome the above difficulties. Addressing these challenges will therefore shape the design of the general SR protocol.

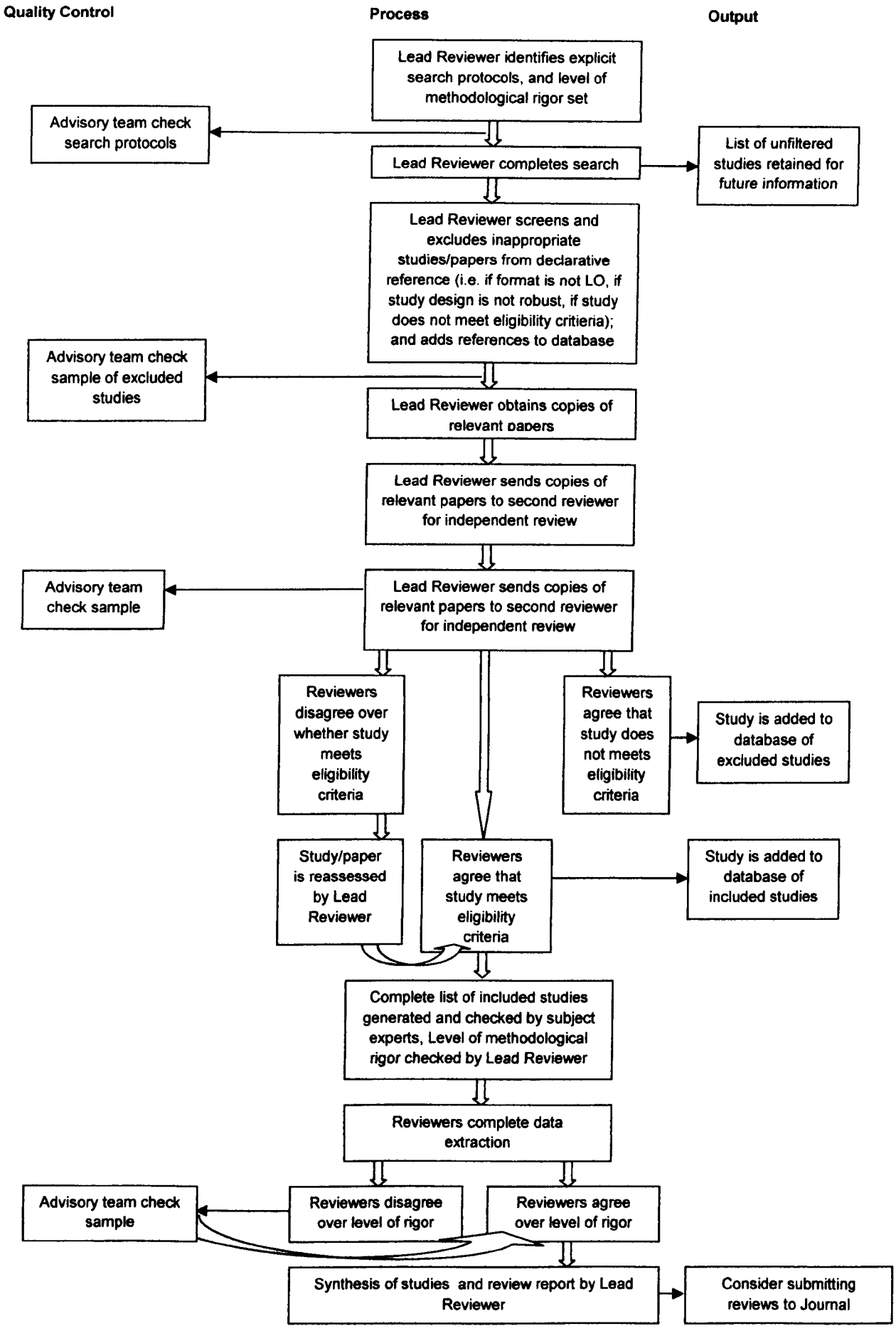
5.3. General Protocol Methodology

No existing protocols were appropriate for the systematic reviews (SRs) as this was uncharted territory. I designed the general protocol in the style of 'Campbell Collaboration' systematic reviews with an amalgamation of the following approaches (each shaping my thinking): Cochrane Group's 'Effective Practice and Organization of Care' (EPOC); Mark Newman & Kate Fleming's use of 'The Campbell Collaboration's protocols' (used by 'The NHS Centre for Reviews and Dissemination'); Sally Collins' 'Cochrane Review Advice'; Instructional Design Theory (outlined in the Introduction chapter); IMS' (2003) 'Learning Design Best Practice and Implementation Guide'; Vernon and Blake's (1993) review studies; Joanna Briggs and BEME Collaboration protocol recommendations; My own thoughts on the literature; the level of 'maturity' in LO research/pedagogy/evaluation; and appropriate level of rigor for healthcare disciplines.

5.3.1. Research Questions, Review Creation and Review Process

As stated, I decided that 3 SRs were needed to establish the effectiveness of all known LO research, pedagogy and evaluation. In each case the overall structure was identical to enable direct comparison and easy translation of findings. This was important as optimum approach/approach combinations could be obtained directly for Phase 2 testing. Each SR was based on this generic protocol and data extraction sheet (Appendix 21) with subject specific data collection questions added to draw out the unique qualities of each (Appendix 22, 23 and 24. N.B. The full generic protocol is too large to be discussed in detail here therefore only the most important points will be outlined). The initial research question was: 'What can previous research tell us about effective LO research types and pedagogy/evaluation approaches?'. The review process is outlined in Figure 11.

Figure 11: Research Process



5.3.2. Data Extraction

Inclusion Criteria, Data extraction, and Methodological Rigor

All papers containing *any* type of LO research, pedagogy, or evaluation were eligible for inclusion in each SR respectively (as it was hoped that generic principles may be later distilled from the findings). This field did not prove to be too wide field thus scope was not limited to health disciplines. I created data extraction sheets (Appendix 25) and scored as follows in an effort to be totally transparent:

DONE - if study participants fulfilled all categories stated;

NOT CLEAR - if insufficient details were given to categorise data (Missing data was discussed with the authors/review leader before further data extraction was undertaken);

NOT DONE - if required information had not been performed by researchers, was not given or was clearly unobtainable.

To ensure that the reviews gathered empirical data with high rigor I designed 'methodological rigor sheets' based on BEME's criteria (Appendix 26). Even if the studies found contained LO research, pedagogy or evaluation in the review population, papers were *only* forwarded for review if they fulfilled the following criteria: i) To be included studies had to obtain the *majority* of ticks in the 'Yes' column (Medical SRs require *all* ticks in the 'yes' column however this is not a medical but an educational review); ii) If the *majority* of ticks were in the 'Don't know' column inclusion was discussed with the Review Leader; iii) If the *majority* of ticks were in the 'No' column studies were excluded.

5.3.3. Participant and Study Characteristics

Anticipated aspects pre-SR were: Sample size, Country, profession, age, subject, academic level of course/training, professional specialty, and study outcomes. Special consideration will be given to the study design type as information as to whether there is a best way to conduct LO research is desired.

5.3.4. Intervention Characteristics

The person/s responsible for conducting the research in each included study will be identified, together with the frequency, duration and intensity of the intervention if possible. QUOROM (2000) suggest examining the follow-up period and how heterogeneity is assessed where possible.

5.3.5. Quantitative Data Synthesis

It was decided that a finding should be labelled 'positive' when group differences show positive changes in learning impact. It is likely that LO studies will use different types of data as outcome measures (e.g. Vernon and Blake's (1993) review includes assessment scores, quality ratings and teaching method preferences) thus analysis will focus on comparing effect differences between the intervention and control where possible. For RCTs/CCTs the baseline and post-intervention differences between study

and control groups will be reported with statistical significance, randomisation/analysis continuity checks where possible. For CBAs the baseline and post-intervention results (with pre-post intervention difference for each outcome) will also be calculated (i.e. post-intervention outcome minus pre-intervention outcome: OE-OC=?), and statistical significance will be included if reported. The selection summary/validity of included papers will be given with simple summary results to illustrate how effect sizes and confidence intervals were calculated (if applicable).

Kirkpatrick methodology (Appendix 27) will be used to estimate the true level of impact of each study included in the SRs thus providing a measurement where effect differences are impossible.

In 2005, I drew up a SR evaluation sheet with guidelines based on the EPPI Review (2002, Appendix 28) knowing that this was –'cutting edge' and intending to publish later. However, in 2006 SCIE published SR evaluation guidelines (Sheet 1 - Appendix 29). When comparing the two, SCIE's guidelines for sheet 1 were very similar to my system. On the one hand this was disappointing not to have published first; on the other it adds further credence to the necessity and value of my work. As there are negligible differences, my system will be compared with SCIE's (and statistical analysis if possible) to see which method is the most appropriate for this type of work.
SCIE also created and recommended using 'Sheet 2'. The full document is inappropriate due to its length/content, thus I created a 'sheet 2 version' with EPPI Review and CRD document-based modifications (Appendix 30).

Meta-analysis

I did not simply want to obtain a 'critical overview' of findings, but a current and rigorous research-based baseline of all available evidence and felt that meta-analysis may aid this process. There is evidence in the literature that 'meta-analysis' is often confused with 'systematic review', however meta-analysis is an *optional SR component* depending upon the level of heterogeneity present to allow meaningful, combination of results (Green 2005, p271). Several definitions are available: 'Meta-analysis' as a critical review discipline that statistically combines previous research's results to summarise evidence on a particular question' (Spector & Thompson 1991); "a mathematical synthesis of the results of two or more primary studies that addressed the same hypothesis in the same way (to) increase the precision of the overall result" (Greenhalgh 1997, p109). This suggests that meta-analysis affords clear, robust and valid combination of results if done well. However, depending upon the number/type of eligible studies found in this research, statistical analysis may be impossible/inappropriate. Optimal precision is desired; therefore a consideration of whether meta-analysis is *appropriate* is needed.

Some authors believe only randomised controlled trials (RCTs) should be subjected to meta-analysis (Spector & Thompson 1991) but few if any RCTs were anticipated. Conversely, others believe that meta-analysis can be performed within SR frameworks (Egger et al 2001). Recent interest regarding education SRs has been described as an 'epidemic of meta-analyses' (Spector & Thompson 1991). However, 'popularity' does not automatically mean 'appropriate for use' hence the deliberation here. As there are directly opposing views, it was decided that a final decision concerning meta-analysis would be taken after reviewing known advantages/disadvantages. Rosenthal & DiMatteo (2001) expound meta-analysis virtues, claiming that most criticism has been based on simple misunderstandings

regarding how quantitative meta-analyses are carried out, and that traditional qualitative narrative reviews are equally guilty. They state that meta-analysis

i) has the ability to 'delineate the research landscape' – this is definitely needed as this thesis's topic is largely unresearched;
ii) keeps statistical significance in perspective. If my research results can be used to develop generic principles this would be invaluable. Undue 'weight' from small studies could be taken into consideration and eliminated where possible;
iii) minimises wasted data and affords the researcher 'intimacy' with summarised data; and
iv) focuses research questions and finds moderator variables.

These two latter points were deemed positive aspects with clear benefits for this research. Spector and Thompson (1991) add two other advantages: "Traditionally, when seeking advice in controversial or novel areas ...scientists have relied heavily on 'informed' editorials or narrative reviews... Meta-analysis can be used to resolve uncertainty when reports... disagree" (p90). They can also be used to combine results of comparable studies to reduce random sampling errors which may skew results for individual studies. In summary, this indicates that meta-analysis may offer a robust, empirical way of resolving many of the problems that educators have when pulling together disparate literature.

However, despite its ready appeal meta-analysis has potential disadvantages:

i) It is currently unknown whether meta-analysis is possible due to the known lack of RCTs, therefore *all* studies should be considered;
ii) Although it is now well-established as a method of reviewing evidence, uncritical use of meta-analysis can be misleading (Thompson 1994). Greenhalgh (1997) recommends ensuring that the methods used are valid/reliable thus emphasis will be placed on rigor, relevance and consistency. Thompson (1994) recommends investigating *heterogeneity sources* and making a distinction between educational and statistical heterogeneity to increase results relevance;
iii) Most meta-analyses performed do not consider individual studies quality (Spector & Thompson 1991). If SR results suggest that 'effectiveness predictions' can be made using certain LO research/pedagogical/evaluation approaches it is crucial that the quality basis on which this is built is known. Quality/rigor sheets should therefore be used and reported on;
iv) Some results are weighted in favour of large studies over small and should instead be weighted in terms of independently assessed quality, derived from predetermined quality criteria: "The pooled estimate can then be adjusted accordingly, or else the quality score used to exclude studies" (Spector & Thompson 1991, p90). For these reasons rigor sheets should determine inclusion/exclusion;
v) Publication bias is a potential problem in all meta-analyses. Spector & Thompson (1991) warn that unpublished papers may contradict meta-analysis findings due to the 'over-representation of published statistically significant studies' (50% of abstracts are never published. Hand-searching of grey literature eradicates this over-representation hence hand-searching should be undertaken. If this is insufficient, funnel plots can be used to represent sample and observed effect sizes, or compare sample size with source population size estimates, (or the number of studies needed to refute the conclusions of the meta-analysis) to estimate publication bias.

Clearly there could be potential difficulties pursuing meta-analysis *unless* strict adherence to the method's rigor and a full quality assessment of weighting/heterogeneity/publication bias is made. SR with meta-analysis would help to pull useful literature together *providing that* i) their objectives, materials and methods are made absolutely clear; ii) they are appropriate for the type of research undertaken; iii) evaluation methods are consistent, robust and transparent regarding conceptual/pedagogical approach; and iv) reviews are reported in a 'transparent' manner so that they can be easily replicated by others.

Despite disadvantages, each appears 'preventable' with careful planning. The above caveats do not work against doing meta-analysis per se and do not outweigh the advantages. Given its 'problem-resolving' advantages I decided to use meta-analysis if appropriate. My working definition for meta-analyses will therefore be 'the critical and statistical appraisal of combined results of all available and eligible studies' (as defined by my SR protocol) in an attempt to summarise, evaluate and 'academically establish' the evidence' concerning the research, pedagogy and evaluation of LOs.

Conducting meta-analyses

There is currently no educational meta-analysis format that is adequate for the task, however several authors have recommendations:

i) The QUOROM format provides a logical system and starting point (Moher D et al, 1999);
ii) Thacker (1990, p91) recommends defining the problem, inclusion criteria, outcome and potential confounding variables; locating research studies; classifying/coding study characteristics; quantitatively measuring study characteristics on a common scale; aggregating findings and comparing with characteristics (i.e. analysis and interpretation); and reporting results;
iii) Chalmers et al (1987) recommend strict adherence to double-blind randomization (to minimise biased intervention allocation); 'transparent' data (to maximise review evaluation/replication); full consideration of missing data (e.g. protocol non-completion); and outcome validation procedures;
iv) Verstraete (2002) recommends strict adherence to a full protocol and studying each trial in detail as data retrieval/selection are crucial regarding meta-analysis validity (p278). However, Verstraete also warns that meta-analyses do not provide the same quality of information as RCTs regarding how effective an intervention is 'if the standards are less stringent'.

Therefore robust, reliable and appropriate meta-analysis should be performed, and to ensure this happens *all* of the above recommendations will be considered. According to (Egger et al 2001) this should result in an unbiased estimate of the intervention in question. (N.B. To comply with PhD regulations I was the Lead Reviewer. To comply with rigorous study selection I appointed/trained a second 'blinded' reviewer who independently evaluated included studies. As both reviewers reviewed all included studies independently before comparing findings biased allocation/inclusion was minimised).

It was possible that this research would require meta-analysis of both qualitative and quantitative data. Green (2005) states that this is possible but not usual for SRs, and warns "Meta-analysis should only be performed when the studies are similar with respect to population, outcome and intervention" (p272). (Statistical meta-analysis will therefore only be performed on sufficiently homogenous data). According

to Thacker (1990), several methods are available: i) quantitative integration of research; ii) 'pooling'; and iii) cluster' analysis.

Quantitative integration of research - This classifies results into statistically significant and non-significant ones in one direction (and opposite direction). It summarises data across studies via vote counting however, "although simple to use voting methods do not take into account the magnitude of effect or sample size" (Thacker 1990, p89). As effect sizes were desired if possible (and because it does not address literature constraints) quantitative integration was rejected.

Pooling - When 'pooling' data (of a single subject from multiple studies), data is combined in a single analysis (Thacker 1990). The point estimate and confidence is calculated to estimate the chance variation. This is followed by a 'pooled average result' across studies if appropriate (Green 2005), and "pictorial representation of the data (forest plot) with a summary measure of effect size (and confidence interval) shown at the bottom of the plot" (Green 2005, p273). The effect size can be used to measure results across studies - i.e. effect size (index of both direction and magnitude) is the difference between two group means divided by the control group standard deviation (Thacker). However, the ability to pool data "is limited by the availability of raw data, variation in study methods, populations under study and statistical considerations" (Thacker 1990, p90). This method depends on copious raw data from SR studies however the amount of available raw data is unknown. Verstraete believes that meta-analysis strengthens the overall evidence and becomes meaningful in practice providing there are no problems of internal or external consistency. However, it would *not* be advisable to recommend an intervention on the sole basis of the meta-analysis - other evidence is needed (Verstraete 2002). This is duly noted.

Cluster Analysis - Regarding this Light & Smith (1971) suggest searching for 'population clusters' and 'explanations and differences'. If these differences can be explained the data can be combined and any statistical variability can be identified. This method had the most potential to be appropriate for this research as it was just dependent upon explainable differences. It also appeared to complement the grounded theory process. Pre-SR it was decided that if meta-analyses were not possible SRs would still be useful as they would show where more evidence or maturity in the field is required. Egger et al (2001, p478) support this: "Systematic reviews are... important to demonstrate areas where the available evidence is simply insufficient and where new trials are required". Despite several advantages of using cluster analysis, the most appropriate method rests upon the actual nature of discovered data.

Statistical Tests

Egger et al (2001) give a useful overview of statistical considerations for meta-analysis (Appendix 31), and Thacker (1990, p91) states that all SRs should include: i) a summary of descriptive statistics and averages of statistics across studies; ii) Calculation of variance (i.e. test for heterogeneity); iii) Correlation of variance by subtracting sampling error; iv) Correction of mean and variance by subtracting sampling error; v) Correction of the mean and variance for study artefacts other than sampling such as measurement error; vi) Comparison of the correction SD to the mean. These will therefore be included in the descriptive statistics given in the results section as appropriate.

Heterogeneity

In SRs, heterogeneity refers to variability/differences between studies' effect estimates. Bandolier (2010) suggests that a distinction should be made between the different types of heterogeneity: i) Statistical (to ensure that perceived effects are not skewed, or under/over-estimated); ii) Methodological (to ensure that methodological rigor is maintained and is not significantly different between studies); and iii) Clinical heterogeneity (Educational heterogeneity will be substituted here as key differences between studies participant characteristics, interventions and/or outcome measures are desired). All three types will be reported on.

Statistical heterogeneity - According to Cochrane there are two ways of identifying the presence of statistical heterogeneity: i) Forest plot with confidence intervals, and ii) χ^2. Further ways include i) the Glance method (Greenhalgh 1997); ii) L'Abbe plot - this explores study inconsistency of studies, event rates and heterogeneous effect estimates (Song 1999, L'Abbe et al 1987, StatsDirect 2010); iii) Funnels plots - test for bias (Egger et al 1997); iv) 'Cochrane Q' test – this compares the chi-square statistic with its degrees of freedom (Cochrane 2010); v) Odds ratio plot – this can be used for both fixed effects and random effects; vi) The I^2 statistic - this describes the percentage of variation across studies that can be attributed to heterogeneity (Higgins and Thompson 2002, Higgins et al 2003). As the nature of the data is unknown each of these will be considered upon completion of data collection as to which is the most appropriate for use.

Methodological Heterogeneity: Higgins et al (2002) advocate 'managing' practical problems to minimise spurious findings from heterogeneity investigations. Unless a large number of studies are available they advise a cautious approach, and warn against using statistical subgroup analyses and meta-regression. They also state: "The appropriateness of using a statistical test for heterogeneity to decide between analysis strategies is suspect" (p1547). This is duly noted.

Educational heterogeneity: CRD (1999), Kober (2001) and Higgins et al (2002) state that heterogeneity issues should be addressed in the protocol and that potential subgroup analyses should be specified a priori to prevent spurious findings from post-hoc subgroup analyses. The potentially important subgroups are therefore identified as disciplines, study level, impact, and outcomes. Oxman & Guyatt (1992) warn against numerous intervention differences within a small trial numbers. This is also noted. Having considered the literature and practice of statistical tests/methods, that 'clustering' may be needed together with the anticipation that a large amount of heterogeneity is possible, the advice of several expert statisticians were sought. Chi2 with/without POISSON was advised (with homogenous studies grouped together using R2 in order to see what studies correlate well together). But, as evidenced above, this *alone* may not be sufficient. Results will be displayed chronologically as outlined above and the estimated intervention effects with confidence intervals will be shown for each study.

5.4. Unique Protocol Methodology

This section outlines the unique parts of the three systematic reviews (added to the generic protocol).

5.4.1. Systematic Review 1: Learning Object Research

The main questions for this review are: What *types* of research studies have been done to date? Which appear to be most effective regarding LO learning? A systematic review on this subject has not been done before so all reported study outcomes will be added to those listed in the data extraction sheets under the title of 'other' in order to be fully inclusive. All unique items added will be discussed /analysed.

5.4.2. Systematic Review 2: Learning Object Pedagogy

The main research question for this review is: What pedagogy is most effective for LO learning? Secondary questions are: What pedagogies have been used? What impact have they had on LO learning? Questions concerning type of paradigm/learning theory and components were added to the generic protocol (Appendix 23). A systematic review on this subject has not been done before so all reported study outcomes will be added to those listed in the data extraction sheets under the title of 'other' in order to be fully inclusive. All unique items added will be fully discussed and analysed.

5.4.3. Systematic Review 3: Learning Object Evaluation

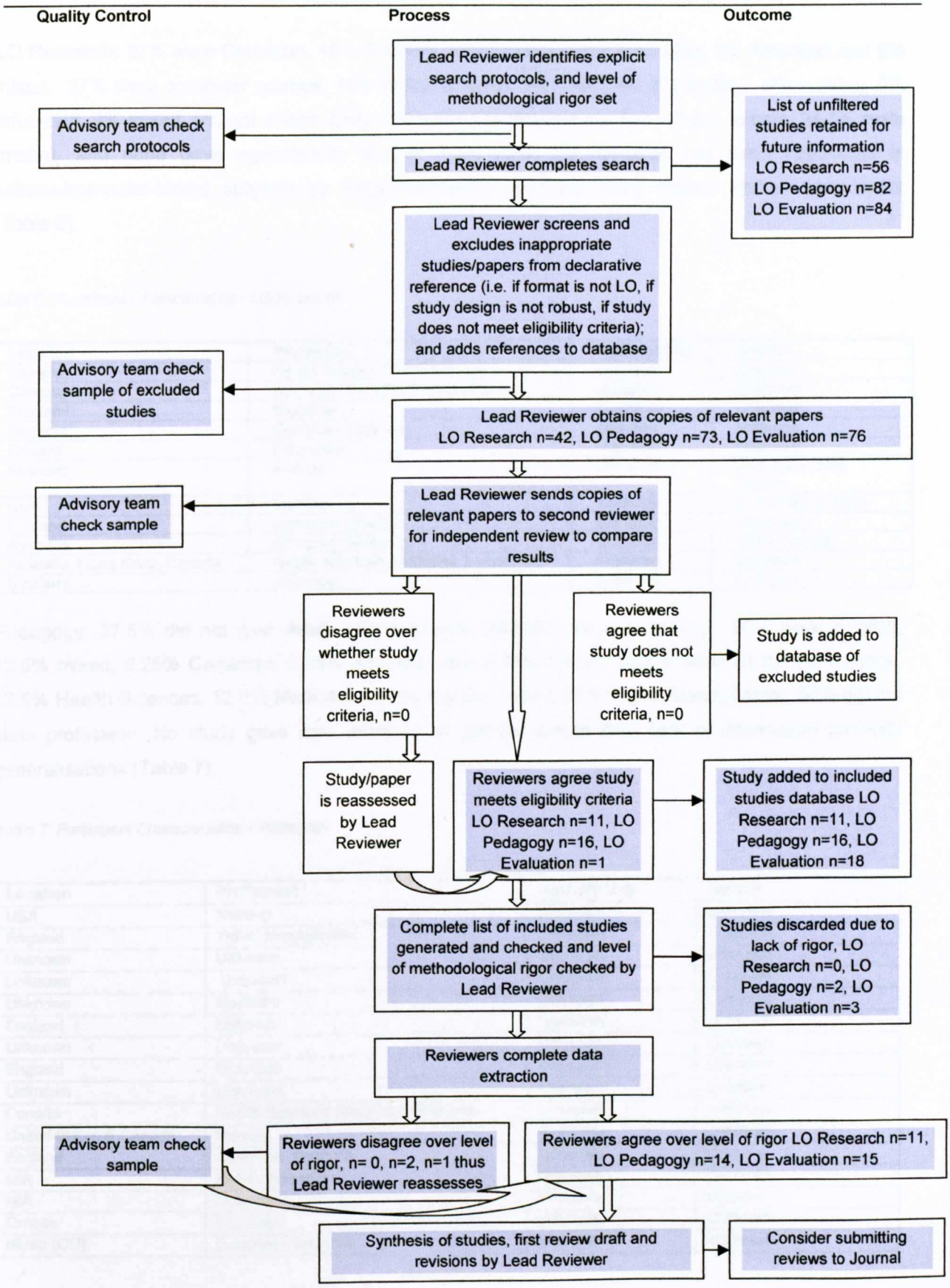
The main research question will be 'What evaluation approaches are most effective in practice?' Secondary concerns are: What outcomes do they measure? What impact do they have on learning? Questions concerning the type of framework and their constituent parts were added to the generic protocol (Appendix 24). Systematic reviews on this subject have not been done before so all reported study outcomes will be added to those listed in the data extraction sheets under the title of 'other' to be fully inclusive. All unique items added will be fully discussed and analysed. Engel's competencies (1991) will be used post-SRs to evaluate how far they can aid the learning of professional competencies. This concludes the protocol items and results will now be discussed.

5.5. Systematic Reviews Results, Discussion and Analysis

All three SRs were undertaken as previously outlined. The generic protocol worked extremely well, elicited detailed data, and translation between findings was uncomplicated. Subject specific components allowed unique properties of LO research, pedagogy and evaluation to be identified and measured. All research questions were fully answered and all major findings will now be shown. The integrated discussion elucidates the major conclusions and significance of my work.

Non-quantitative synthesis is tabulated to allow complementary qualitative assessment of the evidence. Details of the number of papers found for each review and how they were processed are given together with participant, study, and intervention characteristics (with meta-analysis/descriptive statistics). This section also includes the results of all three rigor systems tested (Kirkpatrick, my system and SCIE's system) and the subsequent strengths/weaknesses of each are discussed with recommendations for practice. An in-depth analysis of study type follows. This chapter ends with a protocol efficacy analysis and identifies the rationale for taking top and bottom-performing approaches forward for Phase 2 testing.

Figure 12: The Completed Research Process (The blue depicts the actual route taken through the planned process).



5.5.2. Participant Characteristics

Participant characteristics for each review (LO research, pedagogy and evaluation) are as follows:

LO Research: 27% were Canadian, 18% Scottish, 18% English, 18% Australian, 9% American and 9% mixed. 27% were computer science, 18% health science, 9% medicine, 9% biology, 9% nursing, 9% education, 9% mixed, 9% not stated. Only 27% stated participant gender - 75.6% female, 24.3% male (mean), and none gave age/ethnicity details. Thus most LO research has been conducted in science/computer-based subjects by English-speaking countries using mainly women participants (Table 6).

Table 6: Participant Characteristics - LO Research

Location	Profession	Age/ethnicity	Gender
Canada	Health Science/Pharmacology	Unknown	Unknown
Canada	Business, Creative Writing, Nursing	Unknown	Unknown
Scotland	Medicine	Unknown	Unknown
England	Computer Science	Unknown	Unknown
Canada	Education	Unknown	Unknown
Australia	Biology	Unknown	66% Male, 34% Female
USA	Nursing	Unknown	27 Female, 2 Male
England	Computer Science	Unknown	Unknown
Australia	Computer Science	Unknown	100% Female
Australia, Hong Kong, Canada	Health sciences, medicine & veterinary	Unknown	Unknown
Scotland	Unknown	Unknown	Unknown

Pedagogy: 37.5% did not give details about location (showing poor reporting). 25% were English, 12.5% mixed, 6.25% Canadian, 6.25% American and 6.25% Greek. 12.5% were Computer Science, 12.5% Health Sciences, 12.5% Medicine, 12.5% Nursing, and 6.25% Travel Management. 50% did not state profession. No study gave age, ethnicity, or gender details thus lack of information prohibits generalisations (Table 7):

Table 7: Participant Characteristics – Pedagogy

Location	Profession	Age/ethnicity	Gender
USA	Nursing	Unknown	Unknown
England	Travel Management	Unknown	Unknown
Unknown	Unknown	Unknown	Unknown
Unknown	Unknown	Unknown	Unknown
Unknown	Medicine	Unknown	Unknown
England	Unknown	Unknown	Unknown
Unknown	Unknown	Unknown	Unknown
England	Unknown	Unknown	Unknown
Unknown	Unknown	Unknown	Unknown
Canada	Health Sciences (Nursing & Medicine	Unknown	Unknown
United States & Canada	Nursing	Unknown	Unknown
England	Computer Sciences	Unknown	Unknown
N/A	Health Sciences	Unknown	Unknown
N/A	Unknown	Unknown	Unknown
Greece	Unknown	Unknown	Unknown
Mixed (OU)	Computer Sciences	Unknown	Unknown

Evaluation: 44.4% did not state location, but 22.2% were English, 16.6% mixed, 5.5% Canadian, 5.5% American, 5.5% Greek (Table 8). 38.8% did not state profession but 11.1% were Computer Science,

Sciences, 11.1% Medicine, 11.1% Nursing, 5.5% Management/Commerce/Business/Applied Science, 5.5% Engineering/ IT/Languages. One study gave age/gender and one gave participant nationality. Approximately half of the evaluation research was conducted by English-speaking countries in the area of computer/health sciences. As so little detail has been given regarding SR participant information, no further analysis can be performed nor useful insights gained. Greater detail would have enabled possible relationship predictions regarding 'intervention impact and age/gender', or identify whether top-performing studies come from certain countries for example.

Table 8: Participant Characteristics – Evaluation

Location	Profession	Age/Ethnicity	Gender
USA	Unknown	Unknown	Unknown
England	Nursing	Unknown	Unknown
Unknown	Computer Science	Unknown	Unknown
Unknown	Management, Commerce, Business, Applied Science	Unknown	Unknown
Unknown	Unknown	Unknown	Unknown
England	Nursing	Unknown	Unknown
Unknown	Computer Science	Mean Age = 19 SD=2.06	3 Female, 26 Male
England	Medicine	Unknown	Unknown
Unknown	Neonatal Medicine	4 Danish, 1 German, 1 Greek, 6 UK	Unknown
Canada	Health Sciences - Specialist Nurses/GP Trainees/Medics	Unknown	Unknown
United States & Canada	Unknown	Unknown	Unknown
England	Nursing	Unknown	Unknown
Unknown	Unknown	Unknown	Unknown
Unknown	Study A = Engineering/It/Languages	Unknown	Unknown
Greece	Health Sciences – Nursing/Medicine	Unknown	Unknown
Mixed (OU)	Unknown	Unknown	Unknown
Unknown	Unknown	Unknown	Unknown
Canada And New Zealand	Unknown	Unknown	Unknown

5.5.3. Study Characteristics

Having outlined the participant characteristics, the study characteristics will now be shown.

LO Research

Summary (Table 9): Total studies (n=11). Several studies (n=5) state large numbers but relatively few of these participants *completed* the research. 6 studies do not state size (but the sample size mean for participants completing the research is 126.5). Locations included UK (4), Canada (3), Australia (2), USA (1), and Mixed (1). Professions/disciplines included Computer Science (3), Mixed (2), Pharmacology (1), Medicine (1), Education (1), Biology (1), Nursing (1) and unknown (1). Outcomes included Understanding Clinical/Biosciences; Viability, Costing, Operation, Pedagogy, LO Use/Creation; Pass Rate; Reduced Tutor Marking/Communications/Tutorials; Attitude; Pass Rate & Retention; N/A; LO Reusability; number of LOs produced. 8 included UG degrees and 4 PG degrees.

Table 9: Study Characteristics - LO Research

SAMPLE SIZE	COUNTRY/ LOCATION	PROFESSION/ DISCIPLINE	EDUCATIONAL LEVEL	AGE	OUTCOMES
Unknown	Toronto & Athabasca, Canada	Health-Related (Science, Pharmacology)	Not Stated	Not Stated	Understanding Clinical Sciences
3 Course Leaders and Staff – No= Unknown	Canada	Business, Creative Writing, Nursing	UG Not Stated	Not Stated	Viability, Costing, Operation, Pedagogy, Lo Use/Creation
215 (124 Completed)	Edinburgh - Scotland	Medicine - Second Yr	UG Bachelor	Not Stated	Unknown
650 (36 Completed)	London, UK	Computer Science	UG Bachelor, PG, MSc, HND	Not Stated	Pass Rate
19 (In 2002) 32 (2003) & 15 (2004) = 66	Ottawa & Toronto - Canada	Education	UG BEd, PG MA Med, PG, PhD	Not Stated	Reduced Tutor Marking/Communications/Tutorials
1300 (457 Completed)	Sydney, Australia	Biology - First Year	UG Bachelor	Not Stated	Understanding Of Biosciences
43 (29 Completed)	USA	Nursing	UG Bachelor	Not Stated	Attitude
600 (47 Completed)	London, UK	Computer Science	UG Bachelor, PG, MSc, HND	Not Stated	Pass Rate & Retention
Unknown	Sydney - Australia	Computer Science	PG Masters	Not Stated	N/A
Unknown	Melbourne - Australia, Hong Kong, Guelph - Canada	Health Sciences, Medicine & Veterinary	UG Degree	Not Stated	LO Reusability
Unknown – 27 Projects	Scotland	Various	H/F.Ed. Not Stated	Not Stated	Number Of RLOs Produced

Pedagogy

Summary (Table 10): Total studies (n=18). Few studies (n=2) state sample size. Locations included UK (4), Greece (1), Canada (1), USA (1), Open University (1) and mixed locations (1). 6 were unknown, 3 non-applicable. Professions/disciplines included Computer Science (2), Health Sciences (2), Mixed (3) Medicine (1), Nursing (2), Unknown (4), and N/A (4). Outcomes included Understanding Clinical Sciences; Viability, Costing, Operation, Pedagogy, LO Use/Creation; Unknown; Pass Rate; Reduced Tutor Marking/Communications/Tutorials; Understanding of Biosciences; Attitude; Pass Rate and Retention; N/A; RLO Reusability; and number of RLOs produced. 3 concerned UG degrees, 1 concerned a PG degree.

Table 10: Study Characteristics – Pedagogy

SAMPLE SIZE	COUNTRY	PROFESSION	EDUCATIONAL LEVEL	AGE	OUTCOMES
24	Indiana, USA	Nursing (Research Course)	Distance Learning - Not Stated	Unknown	Unknown
11	London, UK	Mixed – Including Travel Management, Health Science, Medicine and Veterinary Medicine	UG Bachelor	Unknown	RLO Reusability
N/A	N/A	Unknown	Unknown	Unknown	Unknown
Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
N/A	Unknown	Medicine	UG Bachelor	Unknown	Learner-Led Targets
N/A	London, UK	N/A	N/A	N/A	N/A
N/A	Unknown	N/A	N/A	N/A	Gate Principles
N/A	London, UK	N/A	N/A	N/A	N/A

SAMPLE SIZE	COUNTRY	PROFESSION	EDUCATIONAL LEVEL	AGE	OUTCOMES
Unknown	Unknown	N/A	Unknown	Unknown	N/A
N/A	*Toronto* & Athabasca, Canada	*Health Sciences* - Nursing & Medicine	UG Not Stated	N/A	N/A
Unknown	USA/ Canada	Nursing	Unknown	Unknown	Web Curriculum Development/Assess Tutor Experience
N/A	London ,UK	Computer Sciences	UG Bachelor		Unknown
N/A	N/A	Health Sciences	N/A		Unknown
N/A	N/A	Mixed	N/A	Unknown	Unknown
N/A	Different Greek Bodies - ITI & Certh Athens	Unknown	N/A	Unknown	Unknown
Unknown	Open University	Computer Sciences	PG Masters Degree	Unknown	Unknown
Unknown	Unknown	Mixed	Unknown	Unknown	Unknown
Unknown	Unknown	Unknown	Unknown	Unknown	Number of RLOs Produced

Evaluation

Summary (Table 11): Total studies (n=20), 4 have no useful study characteristics therefore the summary will be based on the 16 studies that describe them. One study states large numbers but few participants completed the research. Sample size mean is 129.25 (completed research). 4 studies do not state size. Locations included UK (3), unknown (3), N/A (1), Australia (2), USA (3), mixed (4). Professions/disciplines included Computer Science (2), Mixed (4), Medicine (3), Nursing (2), N/A (3), and unknown (2). Education level included UG (7), HE (4), PG (2), Not Stated (3). Outcomes included: Ability to meet E-Learning demand (Learner-Led Targets, Reduced Time); Sharing of Expertise; Effectiveness of HE Computer Learning via Literature Review and Meta-Analysis; Student Effect Size, Effect of Simulation on Learning; Searching/Use of Formative Evaluation Design/Practice; Student Learning, Community and Economic Exchange. 7 concern UG courses and 5 HE courses.

Table 11: Study Characteristics – Evaluation

SAMPLE SIZE	COUNTRY	PROFESSION	EDUCATIONAL LEVEL	AGE	OUTCOMES
175 (Only 42 Did All 4 Surveys)	Australia	Computer Science	UG Bachelor Elective Module?	Unknown	Unknown
68	Sydney, Australia	Management, Commerce, Business, Applied Science	UG Bachelor - Elective Module?	Unknown	Unknown
450	USA	Not Stated/ IT	Unknown	Unknown	Ability To Meet E-Learning Demand?
Study 2 = 5 Users	Not Stated	N/A	Unknown	N/A	Unknown
65	England	Nursing	UG - Not Stated	Unknown	Unknown
Not Stated	Not Stated	Nursing N/A	Unknown	N/A	Unknown
29	Southampton UK	Computer Science	UG Bachelor	Mean = 19.00 SD = 2.06	Unknown
36	Southampton UK	Medicine	UG Bachelor	Unknown	Learner-Led Targets, Reduced Time
20 (12 Complete)	Southampton UK 6 Denmark 4 Germany 1 Greece 1	Neonatal Medicine	CPD/CME	Unknown	Sharing Of Expertise
Unknown	Unknown	Specialist Nurses/GP Trainees/Medics	Not Stated - Post Registered	N/A	N/A

SAMPLE SIZE	COUNTRY	PROFESSION	EDUCATIONAL LEVEL	AGE	OUTCOMES
59 Studies	Unknown	Not Stated	HE Not Stated	Unknown	Effectiveness of HE Computer Learning Via Literature R/V & Analysis
Not Stated	Baltimore USA	Nursing	UG Bachelor		Unknown
Not Stated	N/A?	Not Stated	Not Stated	Unknown	Unknown
Not Stated	America (Study A)/Europe/Australia	Study A = Engineering/It/Languages	Study A = Higher Education Not Stated	Unknown	Unknown
779 In 9 Studies	Utah USA	Nursing & Medicine	UG Bachelor? Medical Residents	Unknown	Student Effect Size, Effect Of Simulation On Learning
3 Digital Repositories	(Canada) & (N Zealand)	N/A	Higher Education - Not Stated	N/A	1. Searching, 2. Use, 3. Formative Evaluation, 4. Design Practice, 5. Student Learning, 6. Community, 7. Economic Exchange
3 Digital Repositories	(Canada) & (N Zealand)	N/A	Higher Education - Not Stated	N/A	1. Searching, 2. Use, 3. Formative Evaluation, 4. Design Practice, 5. Student Learning, 6. Community, 7. Economic Exchange

5.5.4. Intervention Characteristics

Both participant and study characteristics have now been outlined therefore the main intervention characteristics will now be shown. The researchers for each included study were identified together with the frequency, duration and intensity of the intervention where possible. QUOROM (2000) suggest examining the follow-up period and heterogeneity assessment however insufficient detail made this impossible.

For LO research, lecturers were the predominant researchers. Only one study had a project leader as researcher. The mean frequency was 1.5 times (where stated) which was performed during one course/semester. Duration is not stated except for one study where it was <10hrs.

For pedagogy, lecturers again were the predominant researchers. Only one study cited 'the institution', and one 'project leader'. Frequency was 'once' (where stated); duration was 'over a module' (up to 8m).

For evaluation, researchers were the predominant researchers (30%) with lecturers close behind (25%). Professors and ISD developers were also stated as researchers. Only one study stated the frequency of intervention (as 4) which was undertaken for 4hrs.

5.5.5. Descriptive statistics and Meta-analysis

Having outlined results characteristics, this section now outlines the descriptive statistics and meta-analysis used for all three systematic reviews (SRs). When undertaking SRs Thacker (1990, p91) recommends, i) a summary of descriptive statistics and averages of statistics across studies; ii) Calculation of variance (i.e. test for heterogeneity); iii) Correlation of variance by subtracting sampling error; iv) Correction of mean and variance by subtracting sampling error; v) Correction of the mean and variance for study artefacts other than sampling such as measurement error; vi) Comparison of the correction SD to the mean. However not all of these were possible or appropriate. Sim & Reid (1999)

recommend confidence interval (CI) inclusion whenever a sample statistic e.g. mean is i) presented as an estimate of the corresponding population parameter; ii) provided in addition to/instead of the results of hypothesis tests with the level of confidence for the CI matched to the level of significance for the hypothesis test (e.g. 95% CI for P=0.05, 99% CI for p=0.01); iii) used to assess the importance of study findings (t-tests can be used but often only exclude a difference of zero and do not allow inferences to be made about other possible values of the population parameter); iv) adjusted if multiple CIs are taken to prevent a Type 1 error; or v) reported from individual studies. They recommend that CIs should be displayed in any meta-analysis). CIs were therefore used in this manner.

Pre-SR it was anticipated that this research may require meta-analysis of both qualitative and quantitative data. In reality, only 2 pedagogical studies were found to have mainly qualitative elements so this was not a problem.

Post-SR, all studies except three had minor differences of opinion between reviewers and these were quickly resolved on discussion (study arbitrators/advisors were unnecessary). It is hoped that my strict adherence and the care taken to select appropriate studies for inclusion is self-evident. As previously stated the result, according to Egger et al (2001), is an unbiased estimate of the chosen phenomena.

For RCTs the main outcomes were gathered. The baseline and post-intervention differences between study and control groups were not possible as studies did not mention control group details. For CBAs the main outcomes were gathered for each study and the baseline/post intervention differences between study groups and pre-post intervention differences for each outcome where possible. The pre-post intervention change was measured. Again, few details on control groups were given. For ITSs the main results of the outcomes were gathered. Where studies had impact this was classed as a positive finding (i.e. where differences in the groups were in the intended direction) and the extent of the impact was explored using SCIE, my system, and Kirkpatrick measurement.

As previously stated the nature of the data decided the meta-analysis format. Quantitative integration was rejected as it did not address constraints in the literature nor did it take into account the magnitude of effect or sample size. These were important if gaps in the literature were to be addressed and the true worth of the approach accurately estimated. Pooling was also rejected as it depended on the availability of raw data which was inadequately reported (group means and control group standard deviations were unavailable). Thus Thacker's 'clustering' was used (1990, p89) using R2 as I felt that sample size, impact, and literature constraints all needed to be addressed. Furthermore, pooling was not appropriate as effect sizes could not be calculated (raw data). However differences found were easily explainable indicating that the data can be combined and statistical variability described. As a small number of studies were included in each SR, statistical subgroup analyses and meta-regression were neither appropriate nor attempted.

The general consensus it that statistical analysis cannot be reliably performed on less than 30 studies. However, Hubble et al (1999) states that when predictive statistics are required (as in this case to predict effective approaches) less studies are needed. Only a small amount of studies were rigorous enough to be included in the LO research, pedagogy and evaluation systematic reviews however the aim is to predict effective approaches for Phase 2. The following results are therefore *predictive only.* When differences are explainable the data can be combined and statistical variability described (Light & Smith, 1971). Differences are usually due to heterogeneity thus statistical, methodological and

educational heterogeneity will be discussed now to ensure that i) perceived effects are not skewed or under/over-estimated; ii) that methodological rigor is maintained and is not significantly different between studies; and iii) that differences between studies regarding key participant characteristics, interventions and/or outcome measures are identified).

Methodological heterogeneity - This was kept to a minimum by using a generic protocol for each SR and ensuring that each item was collected/evaluated in the same way. This was also conducted in this way to afford good parity.

Educational heterogeneity - Thompson (1994) states that when there is a degree of clinical (or in this case educational) heterogeneity, statistical heterogeneity would be expected. Findings therefore suggest that a large degree of educational heterogeneity is not present.

Statistical heterogeneity - StatsDirect (2001) note that it is difficult to determine statistical heterogeneity where methodological homogeneity exists, thus methodology heterogeneity was controlled as outlined above. Many intervention differences were found within a relatively small number of trials. Oxman & Guyatt (1992) and Thompson (1994) believe that large meta-analysis problems occur in these circumstances unless heterogeneity is explained. Fortunately studies were much more homogenous than expected. Where heterogeneity existed, much of the difference arose out of the authors' diverse outcomes desired (I wanted top-performing pedagogies/evaluation approaches *regardless of outcome).* Only sufficiently homogenous studies were combined. Even so, Verstraete (2002) believes that it would *not* be advisable to recommend an intervention on the sole basis of the meta-analysis (other evidence is required and will be provided in Phase 2).

Statistical meta-analysis was performed only on data that was sufficiently homogenous in nature using cluster analysis. However, the *extent* of statistical heterogeneity which can be quantified is more important than its existence (DeSimonian and Laird 1986, Thompson 1994). Thus influences on specific educational differences between studies were considered (rather than relying on a statistical heterogeneity test) resulting in more genuine differences and more relevant conclusions being detected. This was wise. Merely using R^2 (RSQ in Excel) resulted in 'how well studies correlate together' rather than identifying top and bottom-performing approaches in a rigorous manner. R^2 was rejected. Chi^2/'Cochrane Q' (CHIDIST in Excel) was considered next but rejected for similar reasons. Additionally, because of the low number of studies, it has low power to detect heterogeneity (Gavaghan et al 2000, Cochrane 2010). (For the evaluation SR, for example, the percentile values were 7.779 at the 90% level, 9.488 at the 95% level and 13.277 at the 99% level with 4 degrees of freedom). When identifying the presence of statistical heterogeneity L'Abbe and funnel plots were not appropriate and could not be calculated as information on control groups/effect sizes was sparse. Similarly, Odds ratio meta-analysis plots could not be calculated. POISSON just predicted the number of events (e.g. effective approaches) over a specific time period however identifying effective approaches per se was required. Calculating minimum/maximum confidence intervals (NORMINV in excel) circumvented this problem by identifying approaches that performed above the 90% level and below the 10% level. Finally, 'NORMINV' (in excel) was used to plot the normal distribution curve with 10%, 25%, 50%, 75%, 90% (and 95%) marked. It was assumed that because I had performed a SR of all available studies and

the sum of the scores taken that this was the *probable* distribution for the whole population. It is hoped that the reader can clearly see that misinterpretation was avoided by i) considering the most appropriate statistical test; and ii) basing meta-analyses on individual participant data from each study rather than summary results (Stewart & Parmar 1993). After NORMIST was calculated the 'glance method' was used to determine level of heterogeneity. Despite the diverse sources, all CIs overlapped (for all three SRs) indicating that study results were sufficiently homogenous in each case to be combined for meta-analysis (See Figure 13 and 14 for pedagogy examples).

Figure 13: NORMIST Confidence intervals for Pedagogy (My system)

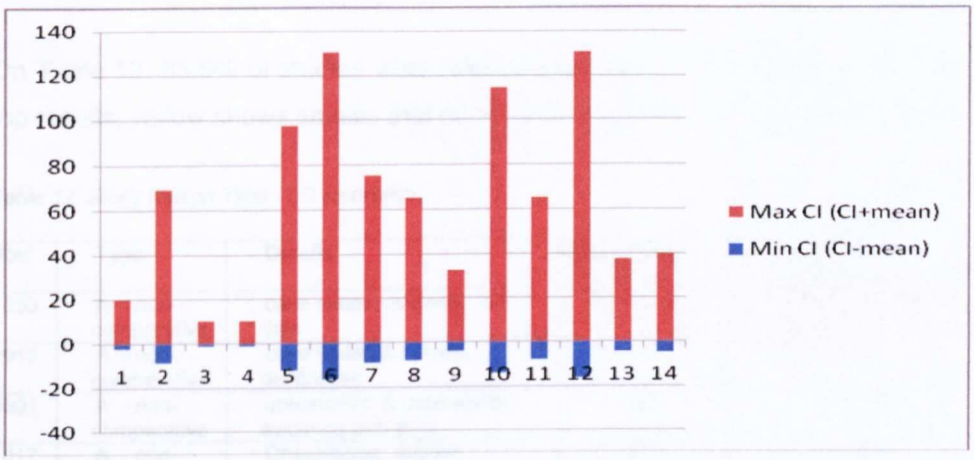
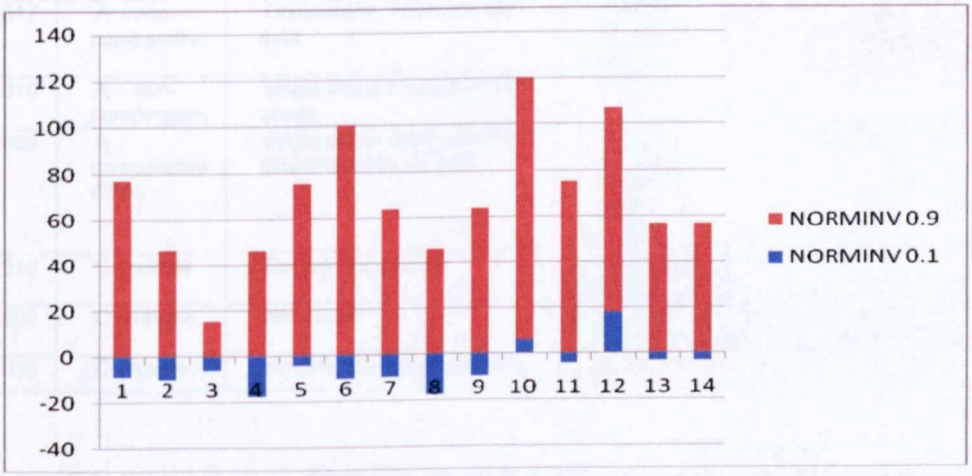


Figure 14: NORMIST Confidence intervals for Pedagogy (SCIE)



To make absolutely sure, this was also checked using χ^2. Some heterogeneity was present at the 90% level for some studies but in each case the p value was above 0.1 indicating that the level was 'negligible'. However, when there are few studies the test is not good at detecting the presence of heterogeneity due to low power thus Thompson's method was followed (a chi^2 statistic/Cochrane Q test having a value equal to its degrees of freedom e.g. number of studies in the meta-analysis minus one). As chi^2 was much lower than the degrees of freedom this supports the evidence indicating that heterogeneity was not present to a large degree and studies were sufficiently homogenous to be combined. Finally, outcome measures were extracted. Positive studies were those that demonstrated a statistically significant change in at least one major outcome measure at the level of $p<0.05$.

5.5.6. Study Design Type

QUORUM's recommendations were followed and study design types are discussed here in detail. As per LO research protocol, design types were important as I wanted to ascertain whether there were optimal ways to undertake LO research. As 'the most popular way' does not always mean 'the most effective way' details were weighed against Kirkpatrick impact and overall evidence level using SCIE and my own scoring systems. Each system was compared to find the best system for LO research, pedagogy and evaluation.

LO Research

On Table 12, 63.6% of studies were non-comparative, 27.2% reviews, 9.1% comparative (Red shows top results, yellow shows studies that came second, and blue shows those that came bottom).

Table 12: Study Design Type - LO Research

Ref	Type	Details	Kirkpatrick impact	Overall evidence level (my design)	Overall evidence level (SCIE)
1030	A - non-comparative	case series, evidence with data	2a - 33%↑	IV = 50%	38.09333333
1013	A - non-comparative	case series & survey, descriptive	1 – 16%	IV = 23.91%	55.05666667
3001	A - non-comparative	observation & case series, evidence with data	1 - 16%	IV = 41.30%	27.97333333
1017	A - non-comparative	Observation, limited evidence with data	1 - 16% examples included	IV = 36.95%	51.19
1004	A - non-comparative	Observation, conceptual	2b - 50% ↑(intended 1)	IV = 30%	52.97333333
3017	A - non-comparative	Observation, evidence with data	1- 16% (intended 2b)	IV = 50%	79.16666667
1016	A - non-comparative	single group interrupted time series	1 - 16%	III = 17.39%	43.75
1008	B - comparative CBAs	single group before & after contemporaneous CBA	1 - 16% overall (2b - 50% for exam results)	III = 17.39%	22.61666667
1018	C - review	Review, conceptual	none, intended 2b	I = 26.5%	72.32
1006	C – review	Not stated	None, intended 2b	I = 14.28%	18.45
1003	C - review	Review, evidence with data	2a - 33%	I = 57.7%	44.64

According to my system, Observation (3017) and case series (1030) came top, both citing evidence with data. A review study (1006) came bottom. According to Kirkpatrick impact, observation (1004) came top followed by case series (1030) and review (1003). A review came bottom (1006), one study gained higher-than-intended impact, and three gained lower-than-intended. This may highlights that it is not just research design that affects ultimate impact. However, the same studies in my system only score at IV level (and have mid-range scores with both this system and SCIE). According to SCIE (if evidence is measured at Level I) the best studies are 3017 (non-comparative, observation), 1018 (review), and 1013 (non-comparative). Review came bottom (1006).

As non-comparative studies appear to be 'top-performing' and reviews 'bottom-performing' in all three systems despite percentage differences this would suggests that non-comparative studies are a good

method for LO research. However, 1003 scores higher than the other reviews which 'breaks the pattern'. On deeper examination it was obvious that this was due to citing data with evidence which accounted for the difference. This finding, combined with the fact that Kirkpatrick's method measures 'impact' rather than 'rigor' per se would suggest that my system and SCIE's system are possibly more rigorous than Kirkpatrick in measuring rigor as the amount of data included seems to affect the Kirkpatrick score. No such limitation was noted for either my or SCIE's system. Thus, it can be said that all systems concur on top/bottom placings but percentage differences in middle-scoring studies may differ. When looking into how SCIE's score sheet was set up it was obvious that these middle-scoring differences are due to the emphasis placed i.e. Weighting within the scoring of relevance of study findings to review, appropriateness of study design, etc (SCIE) for example. As observation came top in all three systems it should be taken forward for Phase 2 testing as the method of choice, however statistics (given after this section) will either confirm or refute this finding. Once statistics have been done, if an approach is clearly identified as top-performing, this will be taken forward for Phase 2 testing using a null-hypothesis. Although further testing of the method itself is not desired for Phase 2 testing, these findings indicate that observation should be chosen so that less confounding variables present themselves (by using a top-performing research method for LO learning).

Statistical analysis

NORMINV in excel was used to plot the normal distribution curve (with all studies) with 10%, 25%, 50%, 75%, 90% (and 95%). As I had performed a SR of all literature in existence, and the sum of the scores were taken, it is assumed that the distribution for the whole *known* population was estimated.

Table 13: Norminv Comparisons of All Three Systems - LO Research

LEVEL	SCIE (SRES1)	MY SYSTEM	KIRKPATRICK
10%	22.31267156	14.63317701	1.170746971
25%	33.5430942	23.43762219	9.745525637
50%	46.02090909	33.22000000	19.27272727
75%	58.49872398	43.00237781	28.79992891
90%	69.72914662	51.80682299	37.37477057
95%	76.45010252	57.07592903	42.50633746

The maximum rigor confidence interval (p=0.05, 95% level) is 42.50 (Kirkpatrick) 76.45 (SCIE) and 57.07 (my system) therefore anything over these values is both performing well and is statistically significant at the 95% level – i.e. study 1004 (Kirkpatrick), 1003 followed by 3017 (my system), and 3017 (SCIE). 90% is 52.11 (SCIE) and 51.80 (my system). Also, any studies performing under the minimum (10%) confidence level of 22.31 (SCIE), 14.63 (my system) and 1.17 (Kirkpatrick) will be taken forward for null-hypothesis testing in Phase 2. My system and SCIE's concurred on the bottom-performing approach (1006, performing below 10% level) and almost regarding study 3017 (performing just below the 90% level). However, it does not concur on 1018 and 1003. Most results were gave similar placings when using SCIE's and my system. Where they differed my system placed more value on the relevance of the study components than SCIE. My system also placed more value on the impact of the study findings, thus when impact level differed the overall score differed using my system but had no impact when using SCIE's system, 3017 and 1018 are given almost same rating in SCIE but 3017 has 457 participants and 1018 has 2. Because SCIE's system gives both studies the same weight, this

highlights that SCIE's system is not as sensitive as my system to this sort of difference (mine show a 23.5% difference).

To complete the calculation, NORMDIST was used giving a normal distribution (for the mean and SD) of 0.98 (Kirkpatrick), 0.58 (SCIE), and 0.87 (my system). To summarise, the results of the SR meta-analyses (i.e. the best/worst studies according to statistical testing of all 3 systems) were as follows:

Table 14: Comparisons of Three Systems According to Top/Bottom-Performing Approaches - LO Research

	Kirkpatrick placing	**Statistical placing: Kirkpatrick**	**SCIE placing**	**Statistical placing: SCIE**	**My placing**	**Statistical placing: My system**
Top	1004 observation	1004 observation	3017 observation	3017 observation	1003 review	1003 review
2nd	1003 review	1003 review	1018 review	1018 review	3017 observation	3017 observation
3rd	1030 case series	1030 case series	1013 case series	1013 case series	1030 case series	1030 case series
Bottom	1006 review (& 1018 review)	1006 review (& 1018 review)	1006 review	1006 review	1006 review	1006 review

1006 in all three systems was the worst (and additionally 1018 for Kirkpatrick) thus review was confirmed as a badly-performing research type using statistics. All approaches to be tested will be performed using observation. 3017 appears in 4 out of 6 of the systems as top or second. For this reason (and the reasons outlined above) observation will therefore be taken forward for top-performing null hypothesis testing in Phase 2.

LO Pedagogy

On Table 15, 43.75% are reviews, 18.75% non-comparative, 18.75% positional papers, and 12.5% are qualitative. 25% are conceptual, 12.5% commentaries, and 12.5% are descriptive in nature (Red shows top results, yellow shows systems that came second and blue shows those that came bottom).

Table 15: Study Design Type – Pedagogy

Ref	**Type**	**Details**	**Kirkpatrick impact**	**Overall evidence level (my design)**	**Overall evidence level (SCIE)**
0003	A - Non-comparative	Observation - descriptive	Unable to ascertain	IV = 8.69%	34.22333333
0004	A - Non-comparative	case study	1 - 16.6%	IV = 30.43%	18.45
0010	A - Non-comparative	narrative commentary	Unable to ascertain	IV = 4.34%	4.76
0015	C - Review	book review/commentary	Unable to ascertain	III = 4.34% (L2)	14.28333333
0009	C - Review	review conceptual	Unable to ascertain	I = 42.85%	35.71333333
0006	C - Review	review conceptual	Unable to ascertain	I = 57.14%	45.23666667
0016	C - Review	review conceptual	Unable to ascertain	I = 33%	27.38
0008	C - Review	conceptual /analytical review	Unable to ascertain	I = 28.57%	14.28333333
0001	C - Review	Descriptive review	Unable to ascertain	I = 14.28%	27.38
0007	D - Qualitative studies	Unable to ascertain	1 - 16.6%	QUAL = 50%	60.11666667
0005	positional paper/book	Unable to ascertain	Unable to ascertain	I = 28.57%	35.71333333
7001	positional paper/book	Unable to ascertain	Unable to ascertain	I = 57.14%	53.57
7003	project report	Project report	Unable to ascertain	QUAL = 16.6%	26.78333333
7002	project report	Project report	Unable to ascertain	IV = 17.39%	26.78333333

Kirkpatrick method could only evaluate impact for 2 studies (1004 & 1007) therefore no judgement can be made as to which studies have the most pedagogical impact. The best studies according to SCIE are 0007 (qualitative), then 7001 (positional), then 0006 (review). The best studies according to my scoring system (if evidence is measured at Level I) are 7001 and 0006 joint first then 0007. This shows that both my system and SCIE identify these 3 studies as being in the top 2-3 places. The 2 studies that score lower have a high level of unobtainable data indicating that the study design/implementation/ reporting are likely to be at fault rather than the research formats themselves. The research methods are very mixed for the top 3 placings and bottom placings which suggests that no one research method is better than another for LO pedagogical research. It is *how research is performed* that matters.

Statistical analysis

Table 16: NORMINV Comparisons of All Three Systems – Pedagogy

LEVEL	KIRKPATRICK	MY SYSTEM	SCIE
10%	UNKNOWN	14.63317701	10.82928404
25%	UNKNOWN	23.43762219	20.15264852
50%	UNKNOWN	33.22	30.51158333
75%	UNKNOWN	43.00237781	40.87051815
90%	UNKNOWN	51.80682299	50.19388262
95%	UNKNOWN	58.46512056	55.77354098

NORMINV in excel was used as before with 10%, 25% 50% 75%, 90%, and 95% marked (Table 16). When comparing the SCIE scoring system with my system (Table 16), the maximum rigor confidence interval is 55.77 (SCIE) and 58.46 (my system) therefore anything over these values is both performing well and statistically significant at the 95% level. These studies were 0007, followed by 7001, then 0006, then 0009 (SCIE) and 7001/0006 followed by 0007 then 0009 (my system). The same 4 studies appear. Lack of data meant that Kirkpatrick could not be evaluated/compared. For the Phase 2 null hypothesis any studies performing under the minimum confidence 10% level (5.24 SCIE, 5.42 my system) were considered to be bottom-performing. These studies were 0010 with 0015 second to bottom (SCIE) and 0010/0015 (my system). 86.6% of the pedagogy studies varied when SCIE and my system were compared (i.e. 40% were slightly higher, 46.6% slightly lower, and 13.3% the same). Despite this difference both SCIE and my system concurs regarding top/bottom-performing studies. To complete the calculation, NORMDIST was used giving a normal distribution of 0.88 for both systems. To summarise, the results of the SR meta-analyses (i.e. the best and worst studies according to statistical testing of all 3 systems) were as follows:

Table 17: Comparisons of 3 Systems According to Top/Bottom-Performing Approaches – Pedagogy

	Kirkpatrick placing	Statistical Kirkpatrick placing	SCIE placing	Statistical SCIE placing	My placing	Statistical my placing
top	Unable to ascertain	Unable to ascertain	0007 Qualitative	0007 Qualitative	7001 positional paper & 0006 review	7001 positional paper & 0006 review
2nd	Unable to ascertain	Unable to ascertain	7001 positional	7001 positional	0007 Qualitative	0007 Qualitative
3rd	Unable to ascertain	Unable to ascertain	0006 review	0006 review	0009 review	0009 review
bottom	Unable to ascertain	Unable to ascertain	0010 commentary	0010 commentary	0010 commentary & 1015 review	0010 commentary & 1015 review

This shows that commentary-type studies perform badly when used for LO pedagogy. Reviews look as though they should be good for pedagogy however reviews are also middle and bottom-scoring when looking at all the studies. This again indicates that no one research method is better than another.

Table 18 shows Chickering and Gamson's '7 principles for good practice in undergraduate education' as the top-performing approach (my system, 94% level) and third (SCIE, 85% level):

Table 18: Pedagogical Approaches Used

Study Ref	Placing	Author/s of approach used	Approach
0006	Joint 1st (me), 3rd (SCIE)	Chickering & Gamson 1989; Active learning (author not stated)	7 principles of good practice; institutions -change management, overcoming in-house problems, staff training. Students - isolation, technology fear, communication
7001	Joint 1st (me), 2nd (SCIE)	Not stated	Not stated
0007	2nd (me), 1st (SCIE)	Dickelman et al	Concept development theory
0009	3rd (me), 4th (SCIE)	Not stated	Not stated
0005	Joint 6th (me), Joint 4th (SCIE)	Halliday 1973 & 1975 and Hall, Hutchings & White 1995	P104 model and Resource-based learning i) resource layer ii) use layer - design potential networks
0008	Joint 6th (me)	Not stated	Not stated
0003	3rd to last (me) 5th (SCIE)	Billings 2000; Jeffries 2000	Distance learning assessment criteria; Dimension Hyper-learning Model
0015	Joint last (me)	Brown, Collins & Duguid 1989; Bransford et al 1990	Situated learning
0010	Joint last (me), Last (SCIE)	Not stated	Not stated

Dickelman et al's concept development theory scores well also (first using SCIE, above the 95% level; second my system, 88% level). Halliday 1973 & 1975 and Hall, Hutchings & White 1995 is just below the 50% level (my system and SCIE) and is therefore too low to be taken forward (50% chance of giving false results). As 0007, 7001, 0006, and 0009 score well, approaches from these studies will be taken forward for testing. Two do not state the approach used therefore Chickering and Gamson's and Dickelman et al's approaches will therefore be tested in Phase 2. Study 0010's approach was not stated by the author however 0015 was joint bottom-performing and is still under the 10% level. Hence Brown et al's 'Situated learning' should be taken forward for null hypothesis testing in Phase 2.

LO Evaluation

On Table 19, 41.1% were non-comparative studies, 41.1% reviews, and 17.8% comparative studies. 17.8% were action-based, 17.8% were commentaries, 11.7% were historical, 11.7% were conceptual (Red shows top results, yellow shows systems that came second, and blue shows those that came bottom). The remaining 51% were surveys citing data with evidence; case series, before-and-after contemporaneous CBAs, retrospective cohort, and RCT type studies. Kirkpatrick method could only evaluate the impact for 37.5% (6 studies). Of these 1 study had lower-than-intended impact. The best studies according to SCIE are 8004, 8002, 8001, 8008/8019 (all reviews). The best studies according to my system are 8004/8002 (if measured at Level I), 8008, 8001, 8019 (identical to SCIE's) suggesting that the review method may be effective for evaluating LOs despite varying percentages between my

system and SCIE. Statistical analysis was performed as before. (Red is top, Yellow is second, blue is bottom).

Table 19: Study Design Type – Evaluation

Study Ref	Type	Details	Kirkpatrick impact	Overall evidence level (my design)	Overall evidence level (SCIE)
8012	A - non-comparative	case series (4) commentary	Unable to ascertain	IV = 26.08% (L1)	16.1175
8010	A - non-comparative	action-based/action research	Unable to ascertain	IV = 23.91% (L1)	40.1275
8013	A - non-comparative	survey- trailing methodology cites evidence with data	Unable to ascertain	IV = 32.6% (L1)	53.015
8015	A - non-comparative	action-based	1 – 16.6% intended 2b	IV = 45.65% (L1)	36.4125
8005	A - non-comparative	action-based descriptive	1 - implied only	IV = 10.86% (L1)	26.5975
8017	A - non-comparative	historical	1 - 16.6%	IV = 4.34% (L1)	8.2275
8006a	A - non-comparative	historical usability	Unable to ascertain	IV = 17.39% (L2)	30.24
8006b	B - comparative	before & after contemporaneous CBA	1 - 16.6%	II = 4.34% (L2)	11.7975
8007	B - comparative	retrospective cohort	1 - 16.6% (2a - 33.3%)	III = 30.76% (L2)	49.8775
8011	B - comparative	RCT	Unable to ascertain	II = 0% (L1)	25
8004	C - review	meta-analysis (Glass 1976)	Unable to ascertain	IV = 85.71% III = 71.42% I = 63.63%	90.1775
8019	C - review	conceptual	Unable to ascertain	IV = 57.14%	57.5875
8002	C - review	commentary Study A = case studies	Unable to ascertain	IV = 85.71% III – 0% (L2)	72.765
8001	C - review	cites evidence with data	1 -16.6%	IV = 71.43% III = 42.85% , I = 27.27% (L2)	60.7125
8008	C - review	commentary	Unable to ascertain	IV = 78.57% (L2)	57.5875

Statistical analysis

As before, NORMINV in excel was used to plot the normal distribution curve at 10%, 25%, 50%, 75%, 90% (and 95%):

Table 20: NORMINV Comparisons of All Three Systems – Evaluation:

LEVEL	KIRKPATRICK	SCIE	MY SYSTEM
10%	UNKNOWN	12.87162322	-0.122209492
25%	UNKNOWN	26.86666225	18.07780069
50%	UNKNOWN	42.41616667	38.29933333
75%	UNKNOWN	57.96567108	58.52086598
90%	UNKNOWN	71.96071011	76.72087616
95%	UNKNOWN	80.33617765	87.61284967

On Table 20, only 6 studies had enough impact to register on the same level of the Kirkpatrick scale therefore no calculations were possible as *SD*=0. The maximum confidence interval for rigor is 80.33 (SCIE) and 87.61 (my system) therefore anything over these values are statistically significant at the 95% level. SCIE found one study (8004) at the 95% level and one at the 91% level (8002). Using my system there were none at the 95% level but 8004 and 8002 performed well at the 94% level. This was

followed by 8001 (above 75% SCIE and my system) then 8019 (just under 75% SCIE and my system). As 8002 and 8004 are reviews (and reviews do not score lower than 57% regardless of system at any time, this appears to suggest again that review-type research performs well for LO evaluation. Studies performing under the minimum confidence level of 12.87 (SCIE) or -0.12 (my system) indicated badly-performing studies. This was 8017 and 8006b for SCIE at below the 10% level, and at the 13% level using my system. Interestingly 8011 (RCT) on both systems did not score well due to lack of basic information given by the author. To complete the calculation, NORMDIST was used (see results below – Table 21) giving a normal distribution of 0.65 (my system) and 0.62 (SCIE).

Table 21: Comparisons of Three Systems According to Top/Bottom-Performing Approaches – Evaluation

Study reference	Placing	Author/s of approach used	Approach
8004	1st (SCIE), Joint 1st (me)	Not stated	Not stated
8002	2nd (SCIE), Joint 1st (me)	Not stated	Not stated
8008	2nd (me), Joint 4th (SCIE)	Not stated	Not stated
8001	3rd (me), 3rd (SCIE)	Not stated	Not stated
8007	5th (me), 4th (SCIE)	Patton 1997 & 2001	utilisation-focused
8019	4th (me), Joint 4th (SCIE)	Patton 1997 & 2001	Utilisation-focused
8006b	Joint 2nd to last (me), 2nd to last (SCIE)	Kara	Layered
8017	Joint 2nd to last (me), Last (SCIE)	Not stated	Not stated

Details for 8004, 8002, 8008, and 8001 were not given by the authors therefore 8007/8019 became the highest-performing approaches (Table 20). This seemed appropriate as 8007 (SCIE) and 8019 (my system) were above the 73% level however it is not ideal as there is a 27% chance that results may be found by chance. Patton's (1997 & 2001) utilisation-focused evaluation approach will be taken forward for Phase 2 testing and discussed accordingly. The worst will be included in order to provide an approach against which to perform a null hypothesis. Details for 8017 as the worst-performing approach were unavailable therefore 8006b (Kara's layered evaluation approach) will was taken forward for Phase 2 testing. This seemed appropriate as 8006b was below the 10% level (SCIE) and at the 13% level (my system).

Table 22: Evaluation Approaches Used

	Kirkpatrick placing	Statistical Kirkpatrick placing	SCIE placing	Statistical SCIE placing	My placing	Statistical my placing
Top	Unable to ascertain	Unable to ascertain	8004 (review)	8004 (review)	8004/8002 (review)	8004/8002 (review)
2nd	Unable to ascertain	Unable to ascertain	8002 (review)	8002 (review)	8008 (review)	8008 (review)
3rd	Unable to ascertain	Unable to ascertain	8001(review)	8001 (review)	8001(review)	8001(review)
4th	Unable to ascertain	Unable to ascertain	8019/8008 (review)	8019/8008 (review)	8019 (review)	8019 (review)
5th	Unable to ascertain	Unable to ascertain	8007 (retrospective cohort)	8007 (retrospective cohort)	8007 (retrospective cohort)	8007 (retrospective cohort)
2nd to last	Unable to ascertain	Unable to ascertain	8006b (contemporaneous CBA)	8006b (contemporaneous CBA)	8006b (contemporaneous CBA)	8006b (contemporaneous CBA)
Bottom	Unable to ascertain	Unable to ascertain	8017 (historical)	8017 (historical)	8017 (historical)	8017 (historical)

5.6. Directions for the Main (Phase 2) Research

Concerning LO Research type, it was clear that no one method gives better results than others as there is an even spread of non-comparative and review types across high to low categories (Appendix 32). As these systems appeared to indicate that any research method is okay if performed well, the observation, interview and usability format will be used again to ensure parity between Phase 1 and Phase 2 findings. Directions for the main (Phase 2) research according to SCIE (Sheet 1), my system and statistical analysis (NORMINV) has already been shown in section 5.5: i.e. Chickering and Gamson/Dickelman (top) and Brown (Bottom) for pedagogy; and Patton (top) and Kara (bottom) for evaluation. SCIE also recommend using what they call 'Systematic Evaluation Sheet 2' (Appendix 33) when examining study efficacy. It scores studies performance as HIGH, MEDIUM or LOW. Sheet 2 (broken down into its various subsections e.g. rigor) will therefore be examined with regard to LO Pedagogy and Evaluation to see whether this can add anything further.

5.6.1. Rigor

Evaluation - Using Sheet 2 Patton's (1887 & 2001) Utilisation-focused approach scored HIGH, Finne et al's (1995) Trailing Evaluation Methodology MEDIUM/HIGH, and a blend of Patton's (1887/2001) Utilisation-focused approach with Shufflebeam's (1971) CIPP approach scored MEDIUM/HIGH (Shufflebeam calls it a 'participant-oriented' approach but it is in fact a Utilisation-focused approach according to BEME criteria). Conversely, Kara & Giannidis et al's (2001) 'Layered' approach scored LOW (See Appendix 33 for a full list). As this concurs with SCIE (sheet 1), my system and statistical findings these will all be taken forward for Phase 2 null hypothesis testing.

Pedagogy – Using Sheet 2 Chickering & Gamson 1989 and Dickelman et al 1999 scored HIGH whenever these were used (despite occasionally being mixed with additional approaches i.e. Active learning theory –Hutchings & White/ Hall & Halliday 1973/1975; Gagne 1985/1992; or Cognitive developmental - Piaget 1970). Hall and Halliday's Active Learning Theory consistently scored MEDIUM to HIGH whenever these were used (despite additional approaches - full list in Appendix 33). Conversely, Emancipatory Model Theory (Rusby 1979) and Active Learning Theory (Gagne 1985/1992) when used alone both scored LOW. As this concurs with SCIE (sheet 1), my system and statistical findings these will all be taken forward for Phase 2 null hypothesis testing.

5.6.2. Research Design Components

A mix of study design components is shown across HIGH to LOW categories (Appendix 34) thus the even spread shows that no one method gives better results than others. Thus no particular design component will be tested in Phase 2.

5.6.3. Data Sources

Here MCQs, practice exams with feedback, quizzes, and tests all score HIGH. The two studies with 'questionnaire, interview, tracking device' mix score LOW (Appendix 35). No real patterns can otherwise be seen thus MCQs will remain in the LOs.

5.6.4. Stated Expected Outcomes

Generally speaking, studies that had 'strategies to aid learning' as their stated outcome scored HIGH, whilst those that stated 'increased pass rate' scored LOW – perhaps indicating that making changes to curricula to increase pass rate *alone* is not enough i.e. students need to use/understand learning strategies. Those focusing on reusability score MEDIUM (Appendix 36). Therefore no particular outcomes will be tested in Phase 2 but learning strategies may need to be considered.

5.6.5. Educational Approach

Generally-speaking short integrated learning resources scored HIGH and modular non-integrated courses scored LOW (Appendix 37). As Phase 2 research is not part of an integrated learning resource significant effort will be made to ensure the LOs are as integrated as possible for the populations tested. This will take the form of liaising closely with stakeholders and the students own wishes.

5.7. Implications of Systematic Review Findings

5.7.1. Pedagogy

For pedagogy top-scoring approaches had either Chickering & Gamson-type principles or Dickelman et al 1999 (concept development) in common. Piaget (1970) (Cognitive developmental approach) scored fairly high as did Hutchings & White/ Hall and Halliday 1973 &1975 (Active learning). Bottom scoring pedagogy included ALL uses of the following approaches: Learning Object Design and Sequencing Theory (LODAS - Wiley 2000 & 2001; Experiential Learning Theory (Bruner 1975, Santrook 2001); Behavioural Connectivist Learning Theory and Emancipatory approaches – Emancipatory Model Theory (Mayes & Fowler 1999, Taylor 1980, Reese & Overton Elemental model 1970 + Knowles et al 1998 Holistic model, Rusby 1979 Emancipatory Model Theory).
Confusingly several pedagogical approaches and active learning theories scored evenly in high, medium and low categories. Further examination revealed that when constructivism was used within an active-learning context (instead of in experiential, 'trial-and-error' or social contexts) it performed well. This further explains why Boyle's deconstruction method came across pedagogical barriers. Boyle's students first entered text on a trial-and-error basis then viewed their results. Inversely, when active learning uses constructivism for real life experiences of learning (as opposed to simulations of learning) it is also effective. When used with other approaches it is not. This explained why a spread of scores was obtained by active learning – it depends how it is used. Further examination also revealed that when pedagogies were based on specific aims (e.g. higher order thinking) as opposed to resource-based non-linear principals, they did not score well. This adds weight to the discussion regarding Deepwell's (2002) linear attempts at developing pedagogy: non-linear (i.e. iterative) principals may be required for pedagogical approaches to perform well.

Paradigms used for low-performing approaches were generally intuitive but largely descriptive in nature. Paradigms in top-performing approaches, by contrast, showed that pedagogical conditions, context, strategies and consequences were required. Situated learning also had a spread of scores

(Wiley et al 2003/Laurillard 2002 scored Medium/High; Hodges & Sasnett 1993 scored Low/Medium; Vygotsky 1962 and Brown et al 1989-95 scored low). Differences between Wiley/Laurillard's use and the lower-scoring approaches were that top-scoring ones were based on real-life contexts and had greater emphasis on use in practice rather than offering theoretical solutions. Again, different scoring was a result of how they were used.

In summary, when pedagogies have an emphasis on use in practice, guided active learning, authentic constructivism and/or deep thought (regarding embedding any paradigm used within the chosen pedagogical environment), they score highly. A summary of top pedagogical approaches and their attributes are shown in Tables 23 and 24:

Table 23: Top Pedagogical Approaches

Top Pedagogical Approaches
Chickering & Gamson's (1989) 7 Principles For Good Practice In Undergraduate Education
Dickelman et al (1999) – Concept Development Theory

Table 24: Top Pedagogical Attributes

Top Pedagogical Approach Attributes
Active Learning + Constructivitism
Respecting diverse ways of learning
Communicating and encouraging learners to have high expectations of themselves
Resource-based non-linear principles and giving prompt feedback
Intuitive Learning + pedagogical conditions, context, strategies and consequences
Emphasis on use in practice, guided active learning, authentic constructivism and embedded paradigms

5.7.2. Evaluation

Top-scoring approaches were either utilisation-focused, used trailing evaluation or were democratic. General evaluation features included student participation in the assessment/evaluation process, documentation of this, and individual learning capabilities. Bottom-scoring approaches were either collaborative - which placed emphasis on context and were primarily devised for curriculum development purposes (e.g. Cousins et al 2002 Communications Model; Cecezkexmanovic & Webb 2000 Communicative Model of Collaborative Learning; Habermas 1984 Theory of communicative action), or 'layered' placing emphasis on authentic/realistic learning (Kara, Giannidis et al 2001). Surprisingly, authentic/realistic learning performed badly. As top-performing pedagogical approaches had identified these attributes for effective LO learning I had assumed that LO evaluation would be the same. This warranted a deeper look. When comparing top/bottom-scoring approaches there was actually no difference in the Kirkpatrick impact (KI) however differences were clearly seen when looking at rigor, strength of findings etc. This meant that the KI either may not be sensitive enough to measure approaches in the way this thesis has set them out, or cannot be used in this way. When comparing this with the pedagogical findings, pedagogical impact was not considered by the SR studies thus no impact score was possible. This indicates that KI either requires more information, or getting a score on the KI is totally dependent upon researching the specific outcomes in the hierarchy. Greater examination of

specific features of the top-performing evaluation approaches revealed that the focus did not concern the *structure* of the process (as in the bottom scoring/adaptive hypermedia-type approaches) but the systematic nature of the process itself i.e. top-scoring approaches were much more systematic in gathering/processing data. This slight difference in emphasis could account for KI incongruities as the KI did not measure the extent to which changes were systematic. Also, when ignoring the KI, it was very clear which approaches should be taken forward as clear patterns were seen in the study attributes. This adds weight to the argument that KI is not appropriate for use/further use.

In summary, when approaches were either utilisation-focused, used trailing evaluation or were democratic they were successful. A summary of top evaluation approaches and their attributes are shown in Tables 25 and 26:

Table 25: Top Evaluation Approaches

Top Evaluation Approaches
Patton (1997) – Utilisation-focused Evaluation

Table 26: Top Evaluation Attributes

Top Evaluation Approach Attributes
Inclusion or capability of student input and/or evaluation in the evaluation process
Capability of producing individual learning plans for students
Able to gather information systematically within the evaluation process itself

The LOs already had an emphasis on use in practice and contained guided positive feedback throughout. I also had intuitively included building knowledge using real-life examples therefore no modifications were needed on these scores. No specific thought had previously been given to embedding paradigms used within chosen pedagogical environments. For this reason, the following question was added to the interview to be used for Phase 2: 'If the appropriate authorisation and technical specifications were approved and you had resources like these LOs to use freely during work time, what do you think would either help or hinder their use?'

5.8. Evaluation of Protocol Efficacy (Post-SRs)

As the main reasons for undertaking the SRs were i) establishing which research, pedagogies and evaluation approaches were most effective; ii) establishing a baseline of LO material; *and* iii) hypothesis generation, this method was deemed 'fit for purpose'. The design of the general protocol appeared to be solid in that no major re-writes were necessary: several new categories were added in the pilot but nothing else was required. The efficacy of the general protocol design will now be discussed in detail.

5.8.1. Research Questions/Review Creation

The overall SR structure proved to be effective as it allowed for easy comparison without further necessary 'transformations'. Inclusion/exclusion criteria, review population and methods were appropriate as they afforded unproblematic analysis. The review process was followed. Individual

reviewer decisions were not checked in detail by the advisory team as no decisions needed advisory team arbitration. Where there was disagreement over levels/categories, discussion between the Lead reviewer (myself) and the second reviewer (PhD supervisor) achieved consensus in all cases. For the LO research review a list of completed studies was sent to subject/field experts in an advisory capacity. One suggestion of a missed study was quickly rectified but no further suggestions were forthcoming. I have performed several literature searches since SR completion to update appropriate studies included but to date no further studies have been found indicating that all appropriate studies are included. The QUORUM flowchart provided a clear method of reporting the overall number of studies and what happened to them at each stage of the process.

5.8.2. Data Extraction

Pre-SR it was anticipated that the minimum methodological inclusion scoring criteria across all study designs would be categorised as DONE, NOT CLEAR or NOT DONE. These categories aided decisions where the decision to include was unclear. It was also anticipated pre-SR that the NOT DONE parts of the study should decide inclusion. When reviewing this post-review, it was clear that NOT DONE parts were crucial in upholding the high level of rigor desired, and did not always mean that the study should be discarded. Studies were discarded if the *majority* of fields within each section fell into this category (e.g. 6+ of 11, 11+ of 20 etc).

Methodological rigor sheets were invaluable: rigor evaluation was unambiguous. The decision to include studies with the *majority* of ticks (as opposed to *all* ticks) in the 'Yes' column proved to be appropriate. There was a real danger that the rigor standard could drop as a result but this was not seen. Rigor was kept high (evidenced by the SRs ability to accurately identify top and bottom-performing approaches when later tested in Phase 2). Studies were excluded where there was a 50% or higher risk of significant bias due to study design weakness. Most studies were only included if there was less than a 25% risk of bias present. It was originally hoped that the accepted level could be set at 'no risk of bias' but this hope was incompatible with the level of academic maturity in the field. Standard EPOC terms gave clear definitions where studies were not easily categorised, where authors had labelled research-type wrongly or left research-type unspecified.

A validity assessment was performed and in all cases no 'masked conditions' were noted (Appendix 42). Authors of included papers were fairly open about their processes/findings, if not a little unclear on occasion. Quality assessment was performed on all studies and a high level of validity was upheld throughout.

5.8.3. Quantitative Data Synthesis

As anticipated a great number of outcomes were identified from included studies, however effect difference/statistical significance could not be calculated for many due to lack of information. Study rigor/impact evaluation therefore often fell to the methodological rigor sheets/SCIE's guidelines and Kirkpatrick's method. Both SCIE and my own system performed really well, whilst Kirkpatrick's method was variable depending upon what emphasis was given and whether studies had varied impact levels

(providing SDs of 0>). The desire to find a rigorous research-based baseline of all available evidence was fulfilled and a full 'geography of the landscape' identified.

Meta-analysis

Sensitivity analysis was used to test search strategies and the robustness of meta-analysis results. Pre-SR it was anticipated that if any studies were 'existing published SRs' they could be used post-SR as 'sensitivity comparisons'. Unfortunately, academic maturity was not sufficiently developed to have such examples; hence search strategy sensitivity was confirmed by contacting LO experts, checking reference lists, and searching the worldwide web. In some instances I asked study authors the reasons for their study choices. Their candid answers were useful at the time, but as the thesis progressed it became clear that their answers contained vital clues regarding how and why their LOs worked/did not work. It is therefore suggested that talking openly and honestly about the underlying reasons for using certain LO pedagogy/evaluation is an important part of developing this field further.
Sensitivity analysis was also used to test meta-analysis results regarding bias. Formal estimates of publication bias effects could not be performed using funnel plots as data was sparse. Whether the SRs were sensitive in accurately measuring top-performing approaches may be best seen by whether they remain top-performing when tested blind and randomised in Phase 2.

The pre-SR debate regarding 'whether meta-analysis was appropriate' was helpful, it aided clear thinking and therefore SR implementation. Effect sizes were not identified but impact, rigor, relevance and strength of findings were. Although large amounts of 'statistical number-crunching' were not possible due to the nature of studies found, basic descriptive and heterogeneity statistics proved more than sufficient regarding the identification of top/bottom-performing approaches. Meta-analysis gave statistical evidence a firm basis on which to rest, enabling rationales for development to be formed. Undue 'weight' from small (or large over small) studies was taken into account and reduced by independently assessing quality using the predetermined quality criteria (SCIE, and my own system). Having a second blinded SR reviewer strengthened this process. Occasionally study reports appeared to superficially disagree but meta-analysis helped resolve this by gathering sufficiently rigorous and homogenous studies together in a format where top/bottom-performing approaches could be identified.

In summary, strict adherence to the method's rigor and full quality assessment of heterogeneity, weighting, and publication bias prevented meta-analysis disadvantages becoming realities. It did this by pulling together information and making their objectives, materials and methods absolutely clear and ensuring that i) they were appropriate for the type of research undertaken; ii) evaluation methods were consistent, robust and transparent regarding conceptual/pedagogical approach; and iii) reviews were reported in a 'transparent' manner so that they can be easily replicated by others.

Statistical tests

There is the possibility of a 'Type I Error' if the following are not stated: i) target size; and ii) whether the analysis stands alone of whether it is part of multiple analyses (p484, Harrigan and Howard Rose). The target size is 'as many as exists' and this analysis is obviously part of the overall analysis (which 1A is a part of) which will be built on as the thesis progresses. Thompson (1994) recommends thoroughly

investigating heterogeneity *sources*, i.e. educational differences between included studies. A type 1 error is also possible if the null hypothesis is wrongly rejected thus every effort will be made to ensure that testing approaches are blinded for Phase 2. Methods chosen to display the results appear suitable.

5.8.4. Unique Protocol Methodology

LO Research: Despite diverse study outcomes the research questions were fully answered. How appropriate/well the research approach is implemented is much more important than the type used.
LO Pedagogy: Despite diverse study outcomes the research questions were fully answered. Exactly why certain pedagogies are better than others is not *fully* understood at this point. It is suggested that despite various work performed in recent years (e.g. Boyle 2002, Boyle/CETL's Positional statement on LO pedagogy/evaluation, JISC's latest pedagogy project) LO pedagogy has not yet *'fully'* matured.
LO Evaluation: Outcomes measured were manifold, yet despite this fact the research questions and secondary concerns were fully answered. Previous LO evaluation work raised the question concerning whether non-traditional approaches were needed. It could be argued that new/non-traditional approaches are not needed as this thesis has found several 'traditional' approaches that apparently work exceedingly well. The effective LO approaches found were not new in themselves but were new *in their LO application*. Initially, I felt that developing pedagogy alongside (instead of seperately to) evaluation may be necessary however pedagogies and evaluation were used (but not in any discernable pairings, nor tested in relation to their interaction within the literature). Thus no attempt will be made to 'pair' them for Phase 2 testing.

Additionally, Engel's competencies (1991) are offered as a means of evaluating how far this review can aid students when learning professional competencies:-

i) Does IDT constitute a robust basis for LO learning? The IDI model makes the LO learning process explicit and fits well with top-performing pedagogies/evaluation approaches. However, despite the resounding success of the IDI model it should be noted that all ISD models have not yet been researched. This therefore constitutes *one operational method with high face validity.* Further research is required in order to estimate the level of rigor that can be assigned to this and other models. As such, no claim can *categorically* be made as to whether this is the *best* model to use. It does however appear to constitute a robust basis for LO learning;
ii) Does present LO theory constitute a robust basis for LO learning? Clearly the monolithic model and Lego-block metaphors are now outdated and may be better suited LO reuse rather than intrinsic value studies. As intrinsic value appears to change depending on the pedagogical/evaluation approach used, Wiley's (1999a) molecular metaphor is more apt. Whereas Wiley uses this at the atomic level, I believe this should be viewed at the *sub-atomic level.* When considering the schematic diagram of an atom (i.e. LO), only certain electrons (i.e. pedagogies/evaluation elements) can be added. When viewing these elements, it could be legitimately questioned as to whether it really mattered or whether there was a real difference between adding one as opposed to two (for example) to a LO. However, it is suggested as in the case of the atom, this distinction may make all the difference to the atom's 'stability' (depending on how many electrons the atom has already before further electrons are added). The number of electrons equates to the number of elements within an evaluation/pedagogical approach which either creates educationally stable LOs (i.e. ones

that 'work') or unstable ones (i.e. ones that don't). If the atom (LO) is stable, it can stand alone (stand-alone LOs); if unstable, it requires other elements to make it work. In summary, some LO theory is now outdated because the field has moved on, but other's (like Wiley's) can be *updated*;

iii) Which parts of Adult Learning Theory are particularly useful/effective in LO learning resulting in increased participant performance at: a) Adapting to and participating in change? The scenario created accommodated all types of answers given during personal development of clinical principles, thus it is particularly useful in its ability to 'adapt to' and 'participate in' change. All other top-performing approaches also adapted to and participated in change however this was heavily reliant on the structure or evaluation parts of the LOs (e.g. guided feedback) and *not* the pedagogy. The LOs themselves adapted to all different learning environments without any noticeable affects; b) Dealing with problems and making reasoned decisions in unfamiliar situations? Due to guided feedback learners could make reasoned decisions and modify decisions if incorrect. The scenario then enabled learners to transfer knowledge into unfamiliar situations. The structure was key in helping students problem-solve and make reasoned decisions in unfamiliar situations.

iv) Adopting a more universal or holistic approach? It is suggested that the learners had to adopt a more universal approach when developing their principles for practice during the scenarios.

In summary. Engel's competencies (1991) served overall as a useful means of evaluating how far this review can aid students when learning professional competencies. When this thesis was commenced JISC's publication was not available but was viewed on SR completion. Despite being developed independently, it supports my findings and Engel's view regarding competency acquisition. It implies that several approaches are both effective and good practice in e-learning, two of which are comparable with this thesis' aims: i) The Associative approach: learning is 'acquiring competence' (Learners acquire knowledge by building associations between concepts and gain skills by progressively complex actions from component skills). This focuses on competences, routines of organised activity, progressive difficulty, clear goals/feedback, and individualised pathways matched to learner's prior performance; ii) The Constructive approach – this focuses on learning as 'achieving understanding' (Learners construct ideas and test hypotheses, p13). This has interactive environments for learning, activities that encourage experimentation/ discovery of principles, and supports reflection/evaluation. In each LO organised activities were staged. After each learning point, the activity difficulty is clearly progressive and immediate *guided* feedback is given throughout. As there is free learner choice concerning which parts of learning to access, the chosen pathways seem to take into account learners existing knowledge by default. The LOs provide an interactive learning environment and encourage experimentation/ discovery of principles in the scenario because of how the learning is constructed. Guided reflection is evident throughout. As both of these approaches have similar elements to the LOs it will be interesting to see whether they are identified as top-performing approaches in Phase 2. In summary, this thesis has already provided some of the necessary evidence regarding top-performing approaches potential to aid the learning of professional competencies.

Sensitivity of the Rigor Systems

As stated in Section 5.5, an analysis of the strengths/weaknesses of Kirkpatrick, the SCIE system and my own rigor system will now be performed. Lack of 'in-depth' data often resulted in the Kirkpatrick

system being unable to measure the impact and consequently comparisons could not be made. Kirkpatrick is also susceptible to 'lower-than-expected' and 'higher-than-expected' measurements if authors do not evaluate the impact in practice after deciding the level of desired impact previously, give accurate account of their findings, or adequately describe the outcomes/effects of their interventions. Both my system and SCIE's system are more rigorous than Kirkpatrick in measuring rigor - amount of data included was shown to directly affect the Kirkpatrick score and its measurement (See p84). Both systems are capable of measuring exceptionally high scoring approaches studies at the 95%> level (See p90) and concur with bottom-performing approaches. However, compared to my system SCIE causes middle-scoring approach differences due to the weighting of *relevance* items (see p84, p85), thus SCIE cannot detect small changes in data relevance like my system does. SCIE also causes middle-scoring approach differences due to the weighting of *impact* (see p85), thus SCIE's system cannot detect changes in impact like my system does. On Table 13 my system detected a 23.5% difference due to being sensitive to the number of participants in the studies (457 versus 2 participants). SCIE's system failed to account for this difference in participants giving each the same weighting. In summary, due to the increased sensitivity of my system this will be taken forward for Phase 2.

5.8.5. Conclusion

SR objectives were to discover what approaches are effective for LO learning. Its aims were achieved and knowledge advanced. Meta-analysis was helpful in pulling useful literature together making i) their objectives, materials and methods clear; ii) ensuring studies were appropriate for the type of research undertaken; and iii) ensuring that my evaluation methods were consistent, robust and transparent (transparent reporting will facilitate easy replication by others). Each study's methodological quality and the extent to which its design/conduct could have either biased or prevented systematic errors were evaluated. The SR protocol was well-executed as approach types and heterogeneity were identified, and meta-analyses performed adding weight to the primary outcome results. This research also successfully identified the most sensitive rigor system for this type of work so this will be used for null hypothesis testing in Phase 2.

6. PHASE 1: MAIN RESEARCH - 1C

In this chapter a short discussion is given regarding compatibility and how easily the two opposing approaches (i.e. Grounded theory - Phase 1A, SR findings - Phase 1B) were integrated. Advantages of the mixed method approach outlined Chapter 1 (Polit and Hungler 1999, p260) serve here to illustrate the extent of compatibility between approaches.

Complementarity - Pre-research I had recognised that each of the approaches, if used alone, would have been 'weak' in the respect that neither could fully identify/address both what approaches were effective in LO learning and why. The deliberate design (for 'each to supply the others lack') worked well as top/bottoming-performing approaches within the systematic reviews (SRs) were easily identified. Grounded theory (GT) hypothesis not only unearthed the reasons why approaches were top-performing but also the participants thought processes whilst they were learning. At first it wasn't *always* obvious where SR findings mapped directly to the GT (e.g. the relationship between LO navigation structure and the perception that 'bad' navigation equates to 'learning styles not being catered for'). Thus there was a high degree of complementary results from the outset but full integration was less obvious initially. Consequently, findings at this stage did not readily or fully 'triangulate'. Deeper reflection was undertaken regarding the extent to which findings *should* triangulate. However, as their intended role was to 'fill in the others gaps' and they more than adequately fulfilled this role, this was not considered to be a problem (During Phase 2 testing, participants will be blinded and randomised under a null hypothesis with different populations, disciplines and locations. This initial small 'incongruity' will be revisited later to see whether the deeper reflection given at this point ultimately resulted in pushing the line of enquiry further - see Overall integration chapter).

Enhanced theoretical insights - When looking at SR and GT, their compatibility and integration capability allowed enhanced theoretical insights to be gained on the nature and behaviour of effective pedagogy and evaluation for LO learning. The initial 'stirrings' of enhanced insights have also been elucidated for 'intrinsic LO worth', 'loop vs. branch learning', and 'LO learning' in particular. As further insights are anticipated post-Phase 2, this discussion will be revisited later (Overall integration chapter).

Incrementality - Incremental/linear progress when developing the GT was possible however a small degree of iteration (as expected) was needed to develop the emerging theory (It is expected that iteration will continue as the theory develops). The visual plotting out of the GT on paper (with directional/causal arrows) helped this process enormously.

Enhanced validity - Up to this stage SR and GT's role has been to enhance each other's findings - which it has accomplished. Once GT is developed for both intended Phase 2 sites a clearer picture of whether the mixed methods approach has truly enhanced the validity of results will be seen. For this reason enhanced validity will not be addressed here, but will be discussed in the Overall integration chapter (methodology section) to determine to what extent the integration of SR and GT approaches have been important in enhancing the measurement and validation of important LO constructs.

Understanding relationships - No phenomena that mattered to the participants appeared to have defied identification/explanation thus understanding the relationships present was very straightforward once visual mapping was underway.

Theory building - My personal understanding has developed as to the size and shape of 'pedagogy' and 'evaluation' constructs and as such an original theory is beginning to emerge. Chapter 1A uncovered some of the reasons why developing LO pedagogy may have been difficult in the past. It is hoped that increased insight on this issue will be gained during the course of main research testing (Phase 2). The comparison with similar or explanatory literature in 1A illustrated where theory gaps are most prevalent. As this thesis is rooted in 'what is useful in practice' these gaps are also the ones most pertinent to practice. Concerning 1B, it was previously anticipated that two general capabilities were required of the SRs: i) they should identify evaluation and pedagogical *approaches* capable of being tested further (i.e. Phase 2) to validate best existing strategies; and ii) approaches discovered by the SRs should be capable of allowing further approaches/hypotheses to be formulated. Both were accomplished. In summary, the weight of evidence shows that these two opposing styles (SR and GT) were highly compatible despite full integration not being initially obvious. The level of obvious integration grew as the constant comparative method progressed. Mixed method advantages have been fully evidenced thus combining the two methods proved to be a good decision.

Recommendations

It can be clearly seen that the work done up to this point (concerning use of the GT format and research structure) provides a firm basis for Phase 2 testing. Methods will therefore remain the same to allow validation of the pedagogy/evaluation* approaches (if the null hypothesis is rejected) and parity between research phases and populations. Both 1A hypotheses and 1B's top/bottom-performing pedagogy and evaluation approaches to be taken forward for testing will be submitted to a null hypothesis in Phase 2 where each participant will be randomised and blinded to the approaches tested.

* The term 'evaluation approaches' in the recommendation section above are labelled as 'evaluation' as this is the term used by the authors who created them. It is recognised that some disciplines may believe 'assessment' to be a more apt term. However the term 'evaluation' will be used in the next phase of the research (rather than 'assessment') as it relates to *building optimal LO structure* rather than the process of assessment.

7. PHASE 2: MAIN RESEARCH - TESTING HYPOTHESES

This chapter seeks to validate and develop the grounded theory (GT), and test both 1A hypotheses and 1B's top/bottom-performing pedagogy and evaluation approaches by null hypothesis. Each participant was randomised to each approach to be tested, blinded to 1A hypotheses/1B approach, and how approaches 'rank' according to 1B findings.

After the necessary preliminaries, Section 7.2 describes how 1A/GT methodology was replicated in a further Midlands-based sample (same population, different participants). Phase 2 findings are compared with 1A/1B findings and discusses how this further develops the overall GT. Should the resulting grounded theory refute the overall theory thus far, this may indicate that the approaches found cannot be generalised within the Midlands medical population. Section 7.3 replicates 1A methodology using a different sample (location/discipline) to see whether this verifies/refutes the overall GT, and Section 7.4 discusses how Section 7.3 findings add to the overall GT theory. Should 7.3 grounded theory refute the overall theory thus far, this may indicate that the approaches found cannot be generalised to nursing populations.

7.1. Preliminaries

For top/bottom-performing pedagogical and evaluation approaches to be tested, the original LOs required modification to comply with exact structure/components of each approach. Context-free principles for each approach were drawn up (See Appendix 38 for excerpts) and LOs modified accordingly (creating tailor-made LOs). This required participants to undertake constituent parts in the order that each approach requires. Procedural commentaries were therefore devised as an 'aide memoire' (and to ensure that strict adherence to each approach was maintained - e.g. Appendix 39). Rigid compliance meant that some participants would not view all LOs parts. For others it meant following a prescribed path, or undertaking additional summative evaluation. A summative evaluation was therefore created (i.e. prioritising care during a real-life based scenario). At no point was any content altered during modifications.

This time, to ensure demographics were readily available, they were added onto the student survey (Appendix 16). The existing Midlands ethics approval was wide enough to encompass further recruitment from the same population as Chapter 1A (with added demographical data) detailed in Section 7.2. Full NHS ethics approval for Eastern Nursing recruitment and further format testing was granted and is detailed in Section 7.3.

Due to the nursing sample change it was important to ensure that all 'test parties' had exactly the same research experience, so to uncover Eastern nurses LO preferences the same generic questions (LO preferences survey) were given to them to complete using the same method (i.e. their most familiar communication method). They were contacted via posters in their workspace advertising the research. From 110 possible responses, 36 responded (32.7%). This was similar to nurses/medics response rates in the Midlands. As surveys already contained demographical questions approved by ethics direct permission was gained to use these. Mean age was 30, and most participants undertook the survey immediately despite having 4 weeks to complete it. As the two sites featured different disciplines it was possible that categories found may differ slightly. However, exactly the same subjects featured top of

both 'most wanted subjects' lists. Independently, without the nurses' knowledge, the corresponding educational stakeholders were asked what their staff/students most needed for practice development. Despite different reasoning the same topics emerged. As both groups at this stage had similar preferences direct pedagogy/evaluation comparisons would be afforded later. If groups had been diverse the differences rather than approaches may have caused extraneous variables (and thus be measured inadvertently).

7.2. Null Hypothesis Testing (Midlands Sample)

In this section the intention was to test 1A hypotheses and 1B top/bottom performing approaches using a null hypothesis on the Midlands site (i.e. there will be no difference between approaches used). 36 medical students were recruited and 1A methodology replicated in order to verify/refute 1A findings and develop GT. Usability studies (with added demographics) were conducted exactly as in 1A.

7.2.1. Working through the GT stages: Results, Analysis and Discussion

Data Collection: Open coding, Constant Comparative Method and Sample

As in 1A, data was collected but participants were randomised and blinded to each top/bottom-performing pedagogical/evaluation approaches. Participants were read pre-formed instructions according to the pedagogical/evaluation principles for the approaches they had been randomised to. When 1A codes appeared, the same code numbers were used. When different ones occurred, they were numbered in chronological order of appearance. Before commencing Section 7.3 it was anticipated that extra codes would be found as this was a totally new set of participants, and this was *testing* newly applied pedagogies/evaluation instead of *finding* them.

Number of Data Units Needed, Selective Coding, Theoretical Sampling, Saturation and Memos

19 extra codes were discovered and written onto index cards with memos. Memos were sorted according to concepts and written up as before. Again, data was grouped by subject and positive/negative/neutral comments so that trends could be seen easily. All participants' answers (from questionnaires, interviews, field note observations) were entered on the devised paperwork. Because analysis of each approach was needed (n=?), and the group was homogenous (as evidenced by their answers), a minimum of 36 data units were collected to give at least 6-8 data units per approach. Incidents were compared and again concepts were easy to link. Properties were confirmed by further data.

Participants were randomised to the approaches therefore recruitment continued until all approaches had enough participants to fulfil saturation. Coding of all participants was undertaken and continued until no new codes/categories appeared. Categories had major consequences with rationales and links were firmly established. This time no modifications to categories were required. Category differences were checked (minimised) to maximise each category's properties. To confirm these, disproving instances were sought for each core category using further new recruits. No further categories were

noted indicating that categories were confirmed and that theoretical saturation had been reached. Most codes gathered were very similar to 1A codes indicating that sample differences were minimal and a high level of parity existed. Data categorisation was straightforward and major categories had good descriptions suggesting that coding was successful. Categories were drawn directly from participants' verbatim comments and their properties complemented what was happening in practice from participant, stakeholder and my own perspective/s. The GT achieved relevance and explained major variations in learning behaviour and was thus deemed 'of good construction'.

Findings: General

General trends included the following:

- 60% of participants preferred information presented as images/animations, 40% preferred written/text elements (i.e. summarised information and bullet points);
- 89% of participants preferred 2+ elements per screen to help focus learning/prevent boredom;
- 50% of participants preferred test elements, 50% favoured 'any interactive method';
- Minor element changes wanted were distributed between visual (i.e. more pictures/animations – 40%) and written elements (i.e. more labels, learning objectives and definitions);
- All participants felt on-screen navigation was good. No negative comments were noted;
- Comments about diagram content were all positive although some participants (3%) wanted slightly larger images.

Compared with 1A findings, the results are mostly similar.

Findings: Underlying Factors

As coding progressed it became obvious that 1A's 11 core codes were exactly the same codes found here with the exception of 'choice' and 'confidence':

Information overload – When elements have the right speed/structure active learning is increased and overload is decreased due to learners feeling that they have had time to 'digest' the learning.
Time – 'Good' pedagogical/evaluation structure decreases overload and increases understanding/ clinical reasoning speed.
Monotony – Images and test elements (i.e. good pedagogy) increase active learning level.
Interest – Images increase the level of active reflection upon learning.
Attention – Increased interaction increases active learning level.
Control – Clear information and user-friendly navigation helps learners decide where they want to go and what they need to learn (i.e. gives pedagogical direction).
Application of learning – 'Good' pedagogy increases understanding which increases reflection (increasing motivation). 'Good' evaluation increases application of learning due to structure/feedback mechanisms.
Motivation – Good text gives the participant 'something to aim for' (i.e. learning goals).
Participant learning preferences – A good element mix shows learners how their understanding is incorrect increasing the active learning level and highlighting required learning direction. This increases the application of learning which increases good knowledge linkage.

Findings: Main Factors

Good pedagogy

- Develops clinical reasoning quicker
- Explains why something works (i.e. increases understanding)
- Applies things to real-life
- Caters for all learner levels

Good evaluation

- Assesses
- Develops clinical reasoning
- Is done frequently
- Enables learners to self-assess

When comparing these main factors with 1A findings, similar themes are evident despite different language used to describe them (e.g. application of learning, all learner levels, etc). This shows good parity between and within this medical student population. Tight linkage between categories was seen. Verbatim notes were checked with participants who verified that these were exact meanings/reflections of their experience thus data was 'theoretically sensitive'. In short, findings verified 1A grounded theory.

7.2.2. Theory Generation

As before, associations within and between single items that relate to the same category (and all relate separately to the core indices e.g. pedagogy and evaluation) were compared (Table 27):

Table 27: Comparing Associations Within/Between Single Items – Pedagogy

Concept=pedagogy	**Minimum = poor pedagogy**	**Maximum = good pedagogy**
Quality of mixed elements	When text/image elements are poor quality reading speed is impaired by feeling distracted/overloaded	When text/mages are high quality clinical reasoning develops quicker increasing enjoyment, motivation and reflection which increases understanding
Clarity of explanation	Understanding decreases when explanation is unclear	Good pedagogy explains why something works therefore increases understanding by making clinically relevant points clear/making it 'real' for the learner
Clarity of practice /theory links	Does not make practice/theory links clear	Applies theory to real-life providing application and transferral of knowledge
Learning levels	Aimed at only one learning level and may totally 'miss the target' for some learners	Caters for all levels of student

Although expressed differently, the pedagogical themes of 'element mix', 'learning application', 'overload feelings', and 'catering for all learner levels' are common to both this and 1A's sample. 'Good pedagogy giving the learner the ability to self-evaluate' is an identical theme to 1A. In both cases the structure of evaluation appears crucial (Table 28):

Table 28: Comparing Associations Within/Between Single Items – Evaluation

Concept = evaluation	Minimum = poor evaluation	Maximum = good evaluation
Assessment	No assessment is possible	Assesses the learner
Clinical reasoning level	Learner does not progress clinically	Develops clinical reasoning due to increased motivation when feedback language is good
Learner level	Learner level cannot be ascertained	Frequent evaluation ascertains learner level and saves tutor time. All levels of learner are found due to the evaluation's structure
Learning objectives	Learners feel unable to judge their progression	When learning objectives are clear this gives learners the ability to self-evaluate

Relationship clusters *between* different codes were also considered (Table 29). All theoretically relevant relationships were extricated for each core index and their underlying factors:

Table 29: Clusters of Relationships Between Different Codes

Grounded Theory Concept	Maximum=appropriate level	Minimum=not appropriate level
Information	Good structure/timing of elements increase active learning by allowing learners to 'digest' learning	Feelings of overload are increased
Time	Good structure/timing of text, pedagogy and evaluation increase understanding and clinical reasoning speed	Feelings of overload are increased
Monotony	Good pedagogy and evaluation (good quality images/test elements) keep attention and increase active learning	Learning is inactive and lacks reflection
Interest	Good audio/images provoke active learning and increase reflection	Learning is monotonous
Attention	Right level of images/ interaction increases attention therefore increasing active learning	Learning is monotonous
Control	Clear information/user-friendly navigation helps learners decide where they want to go /what they need to learn next	Unclear information leads to a feeling of less control over learning
Application of learning	Good pedagogy increases understanding by increasing reflection/motivation. Good evaluation increases application of learning due to good feedback structure	Lack of evaluation/summarised information decreases application of learning
Motivation	Good text gives learners something to aim for (learner targets)	Bad text distracts learners, their speed of reading is impaired and their desire for feedback increases
Participant learning preferences	Good element mix shows learners what is wrong and how. This increases active learning giving direction to ensuing learning needs thereby increasing application of learning and knowledge linkage	Bad mix increases confusion/lack of application

Feelings of overload and personal learning preferences were linked to increased understanding. *Timing* of pedagogical/evaluation elements appear crucial. Links between images, interest, monotony, and active learning were noted. In short, active learning and learning direction/choices feature more highly in this sample than 1A and more specific examples regarding how components affect pedagogy/evaluation are identified.

7.2.3. Hypotheses

Hypotheses formed by this Midlands group were:

- A good LO element mix results in learning errors being identified, increased awareness of learning direction/needs which allows greater application of learning, linkage of knowledge, and development of clinical reasoning;
- Unclear information/Layout (due to lack of clarity/information) decreases active learning resulting in learners feeling that they do not have the time to 'digest' the learning;
- Clear learning objectives give learners the ability to self-evaluate;
- The right level, structure and timing of elements increase active learning by allowing learners to digest learning resulting in increased speed of reading, understanding, clinical reasoning (which does not distract/overload the learner).

These (Phase 2: 7.2) and 1A hypotheses were then compared. As they were very similar direct integration was possible by plotting the theory visually (red denotes Phase 2, black denotes 1A):

- Mixed LO elements result in *learning errors being identified, increased awareness of learning direction/needs which in turn* increased knowledge, interest and achieves a good 'element-interaction' balance *by allowing a greater application of learning, linkage of knowledge and development of clinical reasoning;*
- Layout that is not in line with participant learning preferences *due to lack of clarity or information decreases active learning resulting in learners feeling that they do not have the time to 'digest' the learning* and section links not being detected;
- When information is of good quality learning objectives are *clear* which enables learners to judge the level of learning attained *by giving them the ability to self-evaluate* resulting in them believing that they have progressed;
- Labels *and the right level, structure and timing of elements* giving the right amount of information result in clarity and simplicity and *increase active learning by allowing the learner to digest the learning* results in *increased speed of reading, understanding, clinical reasoning, and* aids navigation and does not *distract or* overload the participant.

Plotting the theory visually

All Phase 2 categories/comments in 7.2 were added into the theory and plotted on paper. New codes were added with causative arrows as before. Complex relationships and inter-relationships were emerging and early in this process the 1A GT was obviously 'correct'. Phase 2 findings not only fitted into existing 1A theory but give it greater explanatory power by adding detail. Each item without exception verified, consolidated or explained 1A theory further.

7.2.4. Core Categories and Comparative Evidence

In line with Glaser and Strauss's method comparative literature, knowledge and experience was collected to see if the GT was similar to existing evidence (Holloway and Wheeler, 1996 p106). Despite the paucity of comparative literature, the following were found:

- 89% of my participants (compared to 97% in the literature) liked 2+ elements per screen to 'help focus learning and prevent boredom'. This is consistent with the dual-process model of working memory, modality and contiguity principles.
- Feelings of overload found here were specifically linked to the structure, speed and timing of elements. When these were at the 'right' speed for the learner, learners were able to 'digest' learning, thus they were less directly linked to channel overload. This supports Jurden's theory – 'correct' speed enables learning to be retained in the working memory resulting in more active learning. This increases the speed of understanding and clinical reasoning development;
- Timing of pedagogical/evaluation elements appear crucial. Links between images, interest, monotony and active learning were noted.
- Active learning and learning direction/choices feature more highly in this sample than 1A. More precise examples of how the components affect pedagogy/evaluation are identified.
- Good pedagogy (enabling learner self-evaluation) is an identical theme to 1A. In both cases evaluation structure appears crucial in that it appears to facilitate the pedagogy. Thus when pedagogy is 'well-accommodated' reading may be easier resulting in increased speed.
- Steve Krug's (2006) publication 'Don't make me think' encapsulates my findings. Indirectly this suggests that e-learning with *intuitive flow* provide the most effective pedagogies.
- Choice and confidence did not come up as core categories. When looking at the breakdown of year groups for this sample it was clear that it contained a higher percentage of year 2 and above. Perhaps learning confidence becomes less of an issue the further the course progresses.

7.2.5. Missing Literature

The links evidenced between 'Timing', 'learning styles/preferences' and 'control' do not seem to be adequately covered by existing evidence. It will therefore be interesting to see whether greater understanding is gained following Eastern sample testing.

7.2.6. Implications for Quantitative Research

Indications suggest that good linkage between GT and the SRs will be crucial. As the SR protocol was completed before GT analysis there is a risk that these links may be difficult. However, if generic principles are possible it is assumed that links should be relatively obvious providing SRs are conducted well. Glaser & Strauss (1967)'s five prepositions were not fully met (i.e. comparisons with a large number of projects with the same hypotheses but none presently exist). Phase 2 findings do however verify IA findings e.g. multi-modal learning increases knowledge retention, interaction and participant interest supports the 'Mixed elements' finding in particular. Illuminating the confirming/disproving instances and their conditions within the study, and confirming central propositions, variables, and dimensions were found was important - any hypotheses created now may later require this capability if generic principle formation is possible. Also, principles needed to be contextually relevant but allow a level of standardisation to effectively function alongside present educational benchmarks.

7.2.7. Systematic Review Data that has a bearing on the Core Theory Hypotheses

Apart from 1B, no LO SRs exist. SR findings do not dispute/refute GT findings but complement them by explaining some of the reasons why different approaches are effective for LO learning. A small amount of triangulation is noted (e.g. some GT codes suggest potentially effective approach types).

7.2.8. Hypothesis Testing

It was previously mentioned that 1A GT was incomplete until SRs had been conducted and findings compared. As this theory has only been tested and developed using two samples from the same population/location it would be premature to claim that the top performing approaches found in this thesis should be used for *all* LO learning. The GT is therefore not considered 'complete' until further Section 7.3 testing is performed.

7.3. Multi-site 'Null Hypothesis' Testing (Eastern Sample)

In this section the intention is to replicate 1A methodology with the LO modifications of 7.2 above (i.e. repeat usability studies and GT using a different sample/location/discipline to see whether this verifies/refutes/develops the overall GT). Again a null hypothesis was used. A sample of 36 nurses from Eastern England were recruited and usability studies (with added demographics) were conducted as in 1A/Phase 2 (7.2).

7.3.1. Working Through GT Stages: Results, Analysis and Discussion

Data Collection: Open Coding, Constant Comparative Method and Sample

Data collection and open coding were performed as in 1A/Phase 2 (7.2). When the same codes appeared, same code numbers were used. When different codes occurred, they were numbered in chronological order of appearance. As this was a totally new discipline, location, and set of participants it was anticipated that numerous extra codes would be found. In reality, a further 23 codes were needed. These were written onto index cards with memos, data was simplified into positive/negative/neutral comments, and participant's answers entered onto paperwork as before. Again, recruitment/coding continued until all randomised approaches had been covered, no new codes/categories appeared, and categories had major consequences/rationales, links firmly established, category differences 'minimised' to maximise category properties, disproving instances sought and theoretical saturation reached. Again, concepts were easily linked and data collected appeared to be 'theoretically sensitive' (giving accurate meaning to data). Data categorisation was straightforward and instances were confirmed by further data. Most codes gathered from the new usability studies were very similar to the codes in 1A and 7.2 above showing a high level of parity.

Number of Data Units Needed, Selective Coding, Theoretical Sampling, Saturation and Memos

Because subgroup analysis (i.e. each approach) was needed and this group was fairly homogenous (as evidenced by their answers, 36 data units were collected to give a minimum of 6-8 data units per

approach (plus several extra were collected to ensure saturation had taken place). As before, codes were then grouped according to subject and positive/neutral/negative statements. Memos were sorted according to concepts and integrated into categories. No category modifications were required. Coding was successful and the GT explained major variations in behaviour thereby achieving relevance.

Findings: General

General trends for the Eastern group included the following:

- 70% of participants preferred information presented as images/animations, 30% preferred written/text elements (i.e. summarised information, simple terminology, and definitions);
- 80% of participants preferred 2+ per screen to focus learning/prevent boredom;
- 40% of participants preferred test elements, 20% favoured any method that increases knowledge, 20% favoured any interactive method, 20% preferred elements that help application of learning;
- Minor changes wanted in elements were distributed equally between visual (i.e. click to enlarge facility/more detail) and written elements (i.e. more labels/information but less text);
- 54% of participants felt on-screen navigation was 'good throughout', 16% felt it was 'mostly clear', 30% felt navigation was 'unclear at times' due to personal preferences (e.g. they wanted learning to go automatically onto the next page when they got the question right instead of having to click onto it themselves). They also wanted extra buttons (e.g. 'back-to-start') on every page;
- Comments about diagram content were all positive although some participants (5%) wanted slightly larger images.

In each case, participants were told to talk about anything they felt was important during usability testing/observation (usability questions were only asked if not covered by free speech or if absolutely necessary. Overall, only 2 participants required this). Given this free range, a comparison between 1A and Phase 2 (Section 7.2 and 7.3) findings reveals that both are *remarkably* similar.

Findings: Underlying Factors

Excitingly, as coding progressed it once again became obvious that 1A's 11 core codes were exactly the same main codes found here (with the exception of motivation, choice and confidence). Descriptions are slightly different but these are *totally different participants*:

Information overload – mixed elements increases active learning and decreases overload
Time – short elements save time due to LO flexibility
Monotony – good pedagogy increases interaction
Interest – increases interaction and active learning/application of learning
Attention – increased interaction increases active learning/application of learning
Control – intuitive LO structure is preferred
Application of learning – 'good' interaction/images increase enjoyment, interest and understanding
Participant learning preferences – mixed elements allow comparison increasing attention, reflection, and active learning due to focusing the mind and showing the learner what is going on.

1A results were then compared with both Phase 2 sites:

Table 30: Comparisons of Phase Results

	1A: Phase 1	Section 7.2: Phase 2	Section 7.3: Phase 2
Information overload	When elements have the right speed/level of text/information learners' confidence increases and levels of anxiety, confusion and the feeling of being rushed decreases	When elements have the right speed/structure active learning is increased and overload is decreased due to learners feeling that they have time to 'digest' the learning	Mixed elements=active learning and decreased overload
Time	When information and layout is well-timed this increases enjoyment and the feeling that the learning is in line with learning preferences. Actual time taken on the package decreases	The structure of pedagogy /evaluation used decreases overload and increases understanding/ clinical reasoning speed	short elements save time due to LO flexibility and structure (pedagogy and evaluation)
Monotony	Indirect only - through interest/attention categories	Images and test elements (i.e. good pedagogy) increases active learning level	good pedagogy increases interaction
Interest	When interest is increased by varying the stimulus, having a good mix of elements, and level/clarity of information this results in learners having learning expectations fulfilled	images increase level of active reflection upon learning	interaction increases active learning and application of learning
Attention	high quality information and images increases attention by 'varying' the stimulus, and increasing interaction	Increased interaction increases level of active learning	increased interaction increases active learning and application of learning
Application of learning	Good audio increases learner engagement due to stimulus change. When engagement increases learners find it easier to apply knowledge and their knowledge desire is satisfied	Good' pedagogy increases understanding which increases reflection/ motivation. 'Good' evaluation increases application of learning due to LO structure and feedback mechanisms	Good interaction and images increase enjoyment, interest and understanding
Motivation	Good text and information 'hits the right level' and increases motivation to learn	Good text gives participants learning goals i.e. 'something to aim for'	Indirect only via text
Participant learning preferences	Low user control over navigation (timing/layout of learning) increases the learners' feelings that the learning is not in line with their learning preferences	A good mix of elements shows learners how their understanding is incorrect increasing active learning levels and giving direction as to what learning is needed next. This increases learning application which increases good linkage of knowledge	A good elements mix allow comparison increasing attention, reflection, and active learning due to focusing the mind and showing the learner what is going on
Confidence	Good text/font speed increases learner confidence. If pedagogy changes this produces decreased learner confidence	Indirect only	Indirect only
Control	Indirect only – via navigation (ISD) and layout (pedagogy)	Clear information/user-friendly navigation helps the learner decide where they want to go and what they need to learn next (i.e. gives pedagogical direction)	intuitive LO structure is preferred
Choice	Clear navigation increases learner perception of having learning 'choice'.	Indirect only	Indirect only

Findings: Main factors

Good pedagogy

- makes you a better nurse by increasing your understanding
- helps images and levels of interaction increase interest, enjoyment and understanding
- combines scientific knowledge and nursing practice
- has differing content which 'focuses the mind' and keeps attention active
- makes learners prioritise care as in real-life
- helps learners apply what they've learned

Good evaluation

- Makes learners aware of what they have learned
- Is easier and more interesting than trying to retain information from book reading
- Tests/builds knowledge
- Helps apply learning
- Reminds learners of what they have forgotten

A quick comparison shows that these findings further consolidate and verify those of 1A/Section 7.2 despite being a totally different population/location/discipline.

7.3.2. Theory Generation

Again, comparing associations within/between single items relating to the same category were performed:

Table 31: Comparing Associations Within/Between Single Items – Pedagogy Theory Generation

Concept=pedagogy	Minimum = poor pedagogy	Maximum = good pedagogy
Quality of mixed elements	Poor interaction level and images decrease enjoyment/attention (learning is less fun)	Good images/interaction levels increase interest, enjoyment , retention and understanding due to being more active/helping learners link knowledge Differing content 'focuses the mind' and keeps attention active
Clarity of explanation	Understanding decreases when explanation is unclear	'Makes you a better nurse' by increasing understanding
Clarity of practice /theory links	Does not make practice/theory links clear when practice/theory balance is unclear	Combines scientific knowledge and nursing practice and aids application of learning
Learning levels	Need to prioritise care goes unnoticed by learner	Good pedagogy makes you prioritise care as in real-life
Simplicity	When layout is not simple learners perceive this as overload	When there is simple terminology and clear layout learners perceive information as being 'evidenced-based'

Although phrased slightly differently, the pedagogical themes of 'element mix', 'application of learning', 'overload feelings', and 'catering for all learner levels' are common to *all* samples. Here, the Eastern nurses have added also a new category to the other samples – i.e. 'simplicity = evidence-based'. When questioned they could not explain their thinking (i.e. tacit perception). On deliberately cluttered pages participants comments were definitely linked with 'overload'. One participant said 'Oh my goodness!' when intentionally presented with a large block of text. The link between block text and simplicity is self-evident, but the evidence-based link is not.

Table 32: Comparing Associations Within/Between Single Items- Evaluation Theory Generation

Concept = evaluation	Minimum = poor evaluation	Maximum = good evaluation
Assessment	No assessment possible	Makes learners aware of what they have learned
Clinical reasoning level	Learner does not progress clinically	Is easier and more interesting than trying to retain information from book reading
Learner level	Learner level cannot be ascertained	Tests/builds knowledge
Learning objectives	Learners feel unable to judge their progression	Helps apply learning

Good pedagogy (enabling learners to self-evaluate) is an identical theme to 1A. In both cases the structure of evaluation appears crucial. Again, clusters of relationships *between* different codes were also considered. All theoretically relevant relationships were extricated for each core index and their underlying factors:

Table 33: Clusters of Relationships Between Different Codes – Theory Generation

Concept	Maximum=appropriate level	Minimum=not appropriate level
Information	Mixed elements increase active learning and decrease overload	Learners feel unable to use the learning and overload feelings are increased
Time	Short elements save time due to LO flexibility and structure (pedagogy and evaluation)	Learning is more limited to time/place
Monotony	good pedagogy increases interaction	Information hits the learner all at once. This 'puts learners off' and learning becomes less active
Interest	interaction increases active learning and application of learning	Learning is monotonous and the learner feels that it cannot be applied
Attention	increased interaction increases active learning and application of learning	Learning is monotonous and harder to understand due to lack of varied stimuli
Application of learning	Good interaction and images increase enjoyment, interest and understanding	Learners perceive the learning as not concise or catering for learning styles and therefore have more difficulty applying it
Participant learning preferences	A good mix of elements allow comparison and therefore increase attention, reflection, and active learning due to focusing the mind and showing the learner what is going on	Learners cannot visually link concepts which results in increased monotony, boredom and overload
Control	Intuitive structure preferred	More or less buttons are wanted by learners resulting in the feeling of lack of control

Feelings of overload and personal learning preferences were linked to increased attention, reflection and active learning. Timing of pedagogical/evaluation elements appear to either limit the learning to 'one place use' or free it to be used anywhere. Links between interaction, interest, monotony and active learning were noted. Active learning features more highly in this sample than 1A. More specific examples of how components affect pedagogy/evaluation are identified.

7.3.3. Hypotheses

Hypotheses formed by this Eastern group were:

- A good mix of elements increase active learning and decrease overload, monotony and boredom
- A good mix of elements allow comparison therefore increasing attention due to the learner knowing what is going on
- A good mix of elements allow comparison therefore increasing active learning which focuses the mind, provoking increased reflection
- Good learning objectives help apply learning
- Good labels make everything absolutely clear

1A and 2 hypotheses were then integrated (black=1A, red=Phase 2 Section 7.2; blue= Phase 2: Section 7.3,):

- A *good* mix of LO elements result in learning errors identified, increased awareness of learning direction and needs

- A *good* mix of LO elements result in increased active learning, knowledge, interest which achieves a good 'element-interaction' balance by a) allowing a greater application of learning, b) greater linkage of knowledge which develops of clinical reasoning, *c) decreases feelings of overload, monotony and boredom;*
- *A good mix of elements allow comparison and therefore increases active learning which focuses the mind, which in turn provokes increased reflection;*
- Layout that is not in line with participant learning preferences due to lack of clarity or information decreases active learning results in learners feeling that they do not have the time to 'digest' the learning and section links not being detected;
- When information is of good quality learning objectives are clear which enables learners to judge the level of learning attained a) by giving them the ability to self-evaluate resulting in them believing that they have progressed, *b) which increases the application of learning;*
- Labels and the right level, structure and timing of elements giving the right amount of information result in clarity and simplicity and increase active learning by allowing the learner to digest the learning results in increased speed of reading, understanding, clinical reasoning, and aids navigation and does not distract or overload the participant.

This indicates that phase 2 findings not only fit into the existing 1A theory but give it even greater explanatory power by adding more detail.

Plotting the Theory Visually

On paper all categories and comments formed by phase 2 were added into the theory formed in 1A. New codes were added with causative arrows as before. Again it was obvious that initial indications were correct as all items fitted into the existing theory. Each item without exception verified, consolidated or explained 1A theory further. Complex relationships and inter-relationships emerged.

7.3.4. Core Categories Compared with Further Literature to Further Develop the Theory

Existing Evidence and How These Findings Link In

As data was similar to 1A there is little further to add here. 80% of participants compared to 97% in the literature liked 2+ elements per screen. This was similar to the 75% found in preliminary work and 89% in the medics group. This remains consistent with the dual-process model of working memory, modality and contiguity principles.

Compared to 1A, feelings of overload were less directly linked to channel overload here but increased understanding via active learning instead. Again this supports Jurden's theory regarding the learning that is retained in the working memory, and active learning theories in general. Links between interest, monotony, and active learning were stronger in this sample than 1A. Learning *application* rather than learning choices (found in 7.2) were noted here, and interaction links are more prominent. This could possibly be explained by the fact that a larger degree of *experienced* participants formed this sample compared to 1A. The mean age of participants was three years older and may have had greater appreciation for the importance of integrating theory and practice (as this tends to develop with

experience). Enabling the learner to self-evaluate is an identical theme to 1A/Section 7.2 but is linked here to learning objectives *and* layout rather than just learning objectives. It is possible that a higher percentage of visual learners were present in this sample but as this was not specifically measured no comment can be made. In both cases the structure of evaluation appears crucial in that it appears to need the ability to complement the pedagogy in order for the pedagogy to 'work' and for the learning to be applied effectively.

Missing Literature

Links between 'Timing', 'learning styles/preferences' and 'control' in section 7.2 could not be commented upon and no comparative literature was available. In all groups tested section links went undetected when layout was not in line with participant learning preferences. The Section 7.3 sample showed that when LO structure/layout is intuitive the learner experiences greater feelings of control. When the learner feels they know what is going on they also feel that learning preferences have been catered for. This is consistently true of all groups. In Section 7.2 timing of pedagogical/evaluation elements appeared crucial but no detailed rationale was possible. It is now clear that good timing of learning elements allows learners to digest learning quicker thereby developing clinical reasoning quicker giving learners the perception that learning preferences have been fulfilled.

Implications for Further Quantitative Research

Having tested the Eastern nursing group the GT 'holds true' when discipline and location is changed as similar findings were gained. Post-thesis, additional locations and disciplines should be tested to further delineate GT boundaries/limitations. Further testing in this way is beyond the scope of this thesis however it will be performed to establish whether effective LO learning principles/approaches are transferable to other *formats* (e.g. mobile delivery). It will also be performed to see whether changing the delivery format changes pedagogical/evaluation approach effectiveness, and whether generic principles and/or model development are possible.

7.3.5. Hypothesis Testing

It was previously mentioned that 1A GT was considered 'incomplete' until all findings were compared. It could be argued that in an 'over-arching' sense that without testing all topics, learning formats, locations and disciplines (beyond the scope of this thesis) the GT remains incomplete. However, this idea will be revisited after further format testing.

7.3.6. Further Comparisons between 1A and Phase 2

General Trends

It can be seen in Table 34 that when General trends were compared (between 1A and Phase 2 Midlands section 7.2/Phase 2 Eastern Section 7.3) that findings were very similar:

Table 34: General Trends Comparison Between Phases

	1A: Phase 1	Section 7.2: Phase 2	Section 7.3: Phase 2
Prefer images	70%	60%	70%
Prefer text	30%	40%	30%
Prefer 2+ elements	75%	89%	80%
Individual preferences	50% test, 50% any practical application method	50% test, 50% any interactive method	40% test, 20% knowledge, 20% interaction, 20% any practical application method
Changes wanted	50% visual, 50% written	50% visual, 40% written, 10% audio	50% visual, 50% written
Good navigation	50% excellent, 42% mostly good, 8% needs changes	60% excellent, 40% mostly good	54% excellent, 16% mostly good, 30% needs change
Larger diagrams	3%	3%	5%

The larger percentage of section 7.3 participants wanting navigation changes could be attributed to many participants choosing to undertake the research at night during their working shift. This meant they were more distracted than 1A/Section 7.3 participants who all chose to take part at a time/place dedicated purely to the research. Poor lighting may also have raised this percentage and may explain why an increased 2% wanted larger images.

In short, these findings are more likely to be attributable to 'altered learning environment' than 'different characteristics'. When searching the literature, a model based on instructional preference developed by Price et al (1977/1989) was found to show that the learner's response to key stimuli, such as environmental (light, heat), affect their learning preferences (Riding and Rayner 1999, Yusoff 2008). This adds further weight to this theory. The percentage difference regarding 'preferring 2+ elements per page' could be related to participant IQ. Mayer (2001) states that when learners have higher spatial IQ there is greater working memory. The highest IQ group (based on entry criteria) is the Section 7.2 sample (medics) and the lowest is 1A (includes non-professional laypersons). Therefore 'amount of working memory capacity' could explain the differences.

Main Factors

It can be seen that when main factors were compared between 1A and Phase 2: Section 7.2 (Midlands)/Phase 2: Section 7.3 (Eastern) that findings were again similar for both evaluation (Table 35) and pedagogy (Table 36).

Table 35: Main Factor Comparison Between Phases – Evaluation

Evaluation		
1A: Phase 1	**Section 7.2: Phase 2**	**Section 7.3: Phase 2**
Good learning objectives give learners the ability to self-evaluate progression	Enables learners to self-assess	Makes learners aware of what they have learned
Clearer LO layout makes self-evaluation easier	Indirect only	Self-evaluation is easier/more interesting (than trying to retain information from book-reading)
Indirect only	Assesses learners	Tests/builds knowledge
Indirect only	Develops clinical reasoning	Helps apply learning
Indirect only	Is done frequently	Reminds learners of what they have forgotten

Table 36: Main Factor Comparison Between Phases – Pedagogy

Pedagogy		
1A: Phase 1	**Section 7.2: Phase 2**	**Section 7.3: Phase 2**
Good pedagogy uses appropriate learning 'channels' and enables learners to evaluate 'learning worth'	Good pedagogy explains why something works therefore increases understanding by making it 'real' for the learner and clinically relevant points clear	Good pedagogy makes you a better nurse by increasing your understanding
Mixed elements on each page have wider learner appeal (learners believe this encompasses more learning styles)	High quality text/images develop clinical reasoning quicker which increases enjoyment, motivation, and reflection. This increases understanding	Good images/levels of interaction increase interest enjoyment and understanding Differing content 'focuses the mind' and keeps attention active
Good learning objectives enable information to be deemed as 'good' by enabling learners to judge their learning progression. If objectives appear too fast learners feel rushed and forget they are in control of learning speed	Indirect only	Good pedagogy makes you prioritise care as in real-life
Lack of 'revisit' options decrease learners feelings of navigational control and increase perceptions that learning is not in line with learner preferences	Good pedagogy caters for all learner levels	Good pedagogy combines scientific knowledge and nursing practice
Good audio increases application of knowledge/linkage to practice. It raises engagement through varied stimulus and satisfies increased knowledge desire	Good pedagogy applies things to real-life providing application and transferral of knowledge	Good pedagogy helps you apply what you have learned
Good pedagogy increases learner confidence. Inconsistent pedagogy creates 'confidence crisis'	Indirect only	Indirect only

7.4 Developing the Overall Emerging Grounded Theory

This section shows how Section 7.2 and 7.3 findings add to the overall emerging theory. When plotting the final GT on paper it was obvious that hypotheses had some missing elements. All previously developed GT hypotheses (1A, Section 7.2 and 7.3) were traced schematically and placing of categories checked (for validity and optimum placing). Where hypotheses routes had missing boxes this indicated a missing code or link. These were inserted and are shown in green:

- A good mix of LO elements result in *increased knowledge resulting in an ability to judge progression therefore* learning errors are identified, *and there is* increased awareness of learning direction/needs
- A good mix of LO elements result in *increased interaction leading to increased* active learning, knowledge, interest which achieves a good 'element-interaction' balance by a) allowing a greater application of learning, b) greater linkage of knowledge which develops of clinical reasoning, c) decreases feelings of overload;
- A good mix of elements allow comparison and therefore increases active learning which focuses the mind, which in turn provokes increased reflection;
- *Unclear* layout *leads to feelings of lack of control and learning choice resulting in feelings that the learning* is not in line with participant learning preferences due to lack of clarity or information this decreases *the feeling that knowledge desire has been satisfied. This decreases the level of* active

learning resulting in learners feeling that they do not have the time to 'digest' the learning and section links not being detected;

- When information is of good quality learning objectives are clear which enables learners to judge the level of learning attained a) by giving them the ability to self-evaluate resulting in them believing that they have progressed, b) which increases the application of learning;
- Labels and the right level, structure and timing of elements giving the right amount of information result in clarity and simplicity *and shows the learner what is going on. This a)* increases active learning by allowing the learner to digest the learning *without* distraction or *feelings of* overload, resulting in increased speed of reading, understanding, clinical reasoning, or *b) the learner feels in control of the learning resulting in good* navigation.

These six statements represent the six main theories produced, tested and verified by this thesis.

7.4.1. Comparing GT with Other Groups

The emerging theory was not only verified using comparison of subgroups (1A), but also by further participants from the same population (Phase 2 Section 7.2), and one substantive group to another (i.e. a different discipline/location/population - Phase 2 Section 7.3). Glaser & Strauss (1990, p55) state that group comparison provides i) control over conceptual /population generality; and ii) simultaneous maximisation/minimisation of data - vital for developing emerging theory. These have been evidenced by the extent to which the GT has developed. The hypotheses develop the concept and scope of LO learning. More importantly code categories are incredibly similar – indicating their potential to be effectively applied in other areas. It is therefore tentatively suggested that the generality of scope can be legitimately extended. Testing further comparison groups will be done post-thesis to highlight any undiscovered data and establish further sets of conditions under which categories exist. This should hopefully establish the full potential for theoretical prediction. Care will be taken to qualify rather than broaden the categories.

7.4.2. Answering Research Questions

The inital GT research questions were 'How do effective approaches inform theory? How do effective approaches inform practice?'. These will now be discussed.

Informing Theory

How effective approaches inform theory was not easy to answer. Several interwoven factors play a part. Previously I wondered whether inserting page position on each page may have circumvented the need to cater for all learning styles. Initially, this question could not be answered. However at this stage of GT development it was clear that not only may inserting page position circumvent the need to cater for all learning styles, but the GT also explained *why*.

Common sense suggests that page number insertion gives greater layout clarity (and may particularly help learners during night shifts as there would be a constant 'present page reminder' of their overall position within the learning). Furthermore, common sense also suggests that effective approaches

should have clear instructions and failsafe mechanisms whereby learners can easily find their way when lost. This is also supported by the fairly large number of people asking for exactly these things during inital usability testing in preliminery work. Logically, one could assume that having these in place would lead to the perception that instructions are clear, navigation is easy and learners are confident concerning their location within the LO. However, once again learner perceptions were found to be *greater* than these. For example, if the flow of arrows are traced (Appendix 40 - see 7.4 plotted hypotheses) it can be seen that improving layout clarity gives learners increased feelings of greater control and learning choice resulting in the feeling that the learning is in line with personal learning preferences *no matter what those preferences are*. This is an important finding. Thus, effective approaches informed theory yet again by telling the researcher *specifically* what the *'value added' learner perceptions are*, i.e. those not immediately discernable.

To create LOs containing optimum levels of personal preferences' for all learners is akin to providing the impossible. This thesis suggests that it is now possible for pedagogical design not to have to incorporate all learning styles, providing that timing of information and layout is 'good'. It is therefore suggested that trying to incorporate all learning styles into LO learning tackles pedagogical problems from the wrong angle. Instead, learning style inclusion appears to be rendered 'insignificant' by the value-added learner perceptions hence greater emphasis should instead be placed on maximising these value-added principles (and therefore effects) contained with LOs during the design phase. Thus it is the *mixture of factors* working and *interacting together within* effective approaches, which in turn renders implicit value-added learner perceptions explicit. It is *this* process which informs theory.

This may explain why Deepwell's attempts to develop pedagogy did not 'reach their full potential'. The value-added learner perceptions were unknown and therefore there was no way of knowing how well it would perform, or how problems with the pedagogy could be resolved or 'dismantled' once encountered. It also explains why Boyle encountered difficulties using a deconstruction/reconstruction approach. The sum of the parts are obviously greater than the whole. This may also go some way to explaining why the New London Group (2000)/Mill's (2006) 'multi-literacies' pedagogy went further before hitting conceptual barriers. Reconstructing the pedagogy into a linear hierarchy or using it in distinct stages may have adversely affected the mixture/balance of factors within the pedagogy, thereby affecting inter-approach and intra-pedagogy interactions. (This idea will be developed even further during Chapters 8 and 9).

Informing Practice

From a learner perpective, there are several indrect ways that effective approaches inform practice 'Good' (i.e. top-performing) pedagogy encourages the same level of prioritisation of care that is required in real-life practice and shows the learner how something 'works'. This gives the learner increased ability to apply learning which increases attention and engagement with the learning. McGee 2003, P6). States that "Evoking engagement in... learning object design is a challenge; each learner may have different ways they are engaged. Additionally, the learning experiences that are wrapped around, proceed, or follow a learning object interaction may effect the engagement of the learner". The effective approaches found by this thesis allow the learner to apply learning in whatever way they choose. Good pedagogy then increases the level of understanding by combining both knowledge and practice. This helps the learner to link practice and theory, thereby developing clinical practice. Good

pedagogy also increases the level of critical reflection via the consideration of different stimuli. This focuses the learners mind creating greater reflection upon practice. When there is a good mix of elements this increases learners knowledge and the 'active' learning level allowing the learner to digest the learning. Greater learning 'linkage' is achieved resulting in increased clinical reasoning development, which in turn increases learning enjoyment and rentention. Obviously, when clinical reasoning and knowledge retention is present, greater patient care is achieved in practice.

Good evaluation helps the learner self-evaluate which in turn gives them the ability to judge learning progression. This results in greater feelings of control over learning/learning choices leading to 'learning expectations' and 'desire for knowledge' being fulfilled. When the learner is confident about the knowledge gained they are more likely to use it in practice.

7.4.3. Conclusions

This thesis has evidenced GT development based on a collection of primary data, and augmented by secondary data from the literature (and later the systematic reviews). The result is theory grounded in empirical data. Phase 1A aimed to discover what underlying relationships contribute to effective pedagogy and evaluation approaches for LO learning. Phase 1B's aimed to discover what the top-performing pedagogical/evaluation approaches were. Phase 1C outlined how 1A and 1B interconnect. Finally Phase 2 aimed to develop the GT and test all top and bottom pedagogical and evaluation* approaches discovered. All of these aims have been fulfilled adding further knowledge to the field.

As previously mentioned, it was unknown whether generic principles existed or would be found. As each group tested was given 'free reign' (regarding Phase 1 usability comments), and GT codes and findings are very similar it is likely that generic principles exist. This evidence is promising and provides a great basis for further development using different delivery formats.

* As these evaluation approaches have now been identified they can potentially be used for either evaluation or assessment. They refer to the components necessary within LO structure to provide optimally effective evaluation/assessment structure. It should be noted that they have been tested in relation to the structure they provide for optimally effective learning.

8. FURTHER FORMAT TESTING, PRINCIPLES AND MODELS

I decided that further format testing should take place for four reasons: i) to establish the emerging grounded theory further; ii) to test whether generic principles truly exist/can be used for mobile learning; iii) to see whether LO 'model-building' is possible; and iv) to see whether this thesis's approach also 'works' for mobile learning. The LO definition will remain as in the Phase 2 chapter and this chapter will replicate the process already used i.e. Phase 1A: Usability testing (observation, questionnaires and interviews) and Grounded theory (GT) hypothesis production, Phase 1B: Systematic review (SR), Phase 2: Hypothesis testing (N.B. 1B will be integrated with Phase 1C due to words limits). Up to this point in the thesis desktop LOs have been used, however it was wondered whether top-performing approaches would also 'work' using mobile technology (i.e. learning any time/place/anywhere with laptops, PDAs, iPads, and mobile/smart/i-phones). Full NHS ethical approval had been granted for an Eastern Intensive Care Nursing (ICU) population (see Phase 2, Section 7.3 for details) so an ICU worked example will be discussed to see whether generic principles and/or model of LO learning are possible. The following will be illustrated: i) how the 'distilling' process was performed, ii) how each model stage should be performed; and iii) how the model may be incorporated into various settings.

Kukulsha-Hulme and Petit 2007, p1492) define mobile learning as "learning that is not time or space dependent" and note that it can be 'informal, unobtrusive, ubiquitous and disruptive'. Peters (2007) describes it as learning performed on handheld/desktop devices that are portable, interactive, connected and individual. Thus the working definition of mobile learning used here will be '*any* type of learning performed on *any* mobile device unconstrained by time or place'.

8.1. Phase IA: Usability Study and Grounded Theory

8.1.1. Usability Study

1A was replicated with the ICU sample (for mobile use). A usability study was necessary to ensure that mobile learning was appropriate for the required context, and ii) to gain data for hypothesis testing. As no usability questionnaire specifically established for mobile learning existed, the 'desktop' LO questionnaire previously created/used was reviewed, adapted and specific mobile learning research questions incorporated. Once again, to produce a high level of parity/gain further insight, the study also included observation and interview. It was conducted exactly as in Chapter 7 (Phase 2 testing) on randomised/blinded ICU nurses to i) observe how they use devices; ii) ascertain their thought processes; and iii) check that operational features did not introduce confounding variables.
No modifications to the LOs, pedagogical/evaluation approaches, inclusion criteria or randomisation methods were made thus the ONLY change (compared to Phase 2: Section 7.3 testing) was the delivery format (i.e. mobile devices). The same LOs were tested on the same ICU population with different participants blinded to the approach being tested. (N.B Full unit policy for their use was adhered to during testing). Ideally, further testing on *all* of the same groups would have created *full* parity (i.e. medicine, science education and IT backgrounds) however the ethics committee initially failed to grasp this and wished to restrict the research to *student nurses on ICU only (n=2)* as they felt

that the student nurse/medics population was over-saturated. After a face-to-face meeting (stating the case for parity, etc) the committee's understanding increased thus access to all levels of ICU nurses was given n=360+ (but disappointingly no access to medical students). Further format testing will therefore only be compared with Phase 2 Section 7.3 (nursing participants on 2 sites) unless all findings are sufficiently similar for comparisons to be made. As before, recruitment continued until all approaches had at least 5 participants (n=36). Data from each approach discovered during this process was analyzed.

8.1.2. Grounded Theory (GT)

GT was used to develop emergent theory from usability studies. Participants' verbatim statements were taken put through Glaser and Strauss' approach using the constant comparative method. Chapter 1A's process was replicated and performed as before, but due to lack of space *full* findings for 1A's further format testing will not be detailed here (just general trends).

Findings: General Trends

- 60% of participants preferred information presented as images/animations whilst the remaining 40% preferred written/text elements (summarised information, simple terminology/ definitions);
- 89% of participants liked 2 or more elements per screen to help focus learning and prevent boredom
- 50% of participants preferred the inclusion of test elements, the remaining 50% appeared to favour any method that aided the practical application of learning;
- Minor changes wanted in elements were equally distributed between visual and written elements;
- Most participants felt navigation through the learning material on-screen was good and that content was clear and well-organised. Remaining negative comments were minor;
- Comments about content of diagrams were all positive. No participants wanted larger images.

When compared with 1A/Phase 2 *desktop* findings, the results are very similar with the exception that 16% of this mobile sample felt that *extra* links would enhance the learning further. These findings were similar enough to *1A desktop delivery* to warrant continued hypothesis testing (see Phase 2 later in this chapter). When comparing 1A above with Chapter 4's findings (1A) it can be seen that the GT is very similar. These were formed upon learners comments during observation of use indicating that the delivery format does not change the way students think about their learning. However, to be able to form a model of LO learning, I felt that it was necessary to repeat the whole process (1B, Phase 2) to be certain of the results, thus 1B was commenced.

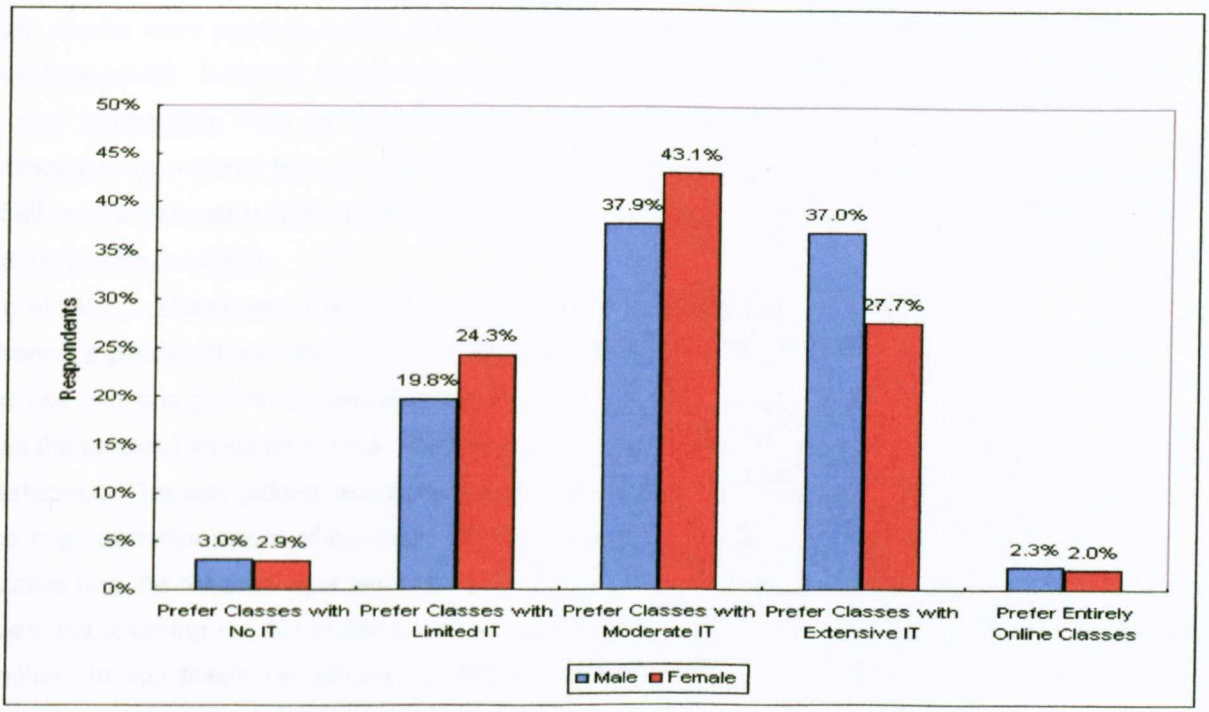
8.2. Phase 1B: Systematic Review

8.2.1. Background

Today's students are described as 'techno-savvy' (Kvavit 2004) or 'digital natives' (Prensky 2001), preferring teamwork and experiential activities (Kvavit 2004). This was evidenced in my preliminary study findings where students had good access/skills and felt confident using computers (97%+). When

Kvavit (2004) surveyed 4,374 students at 13 HEIs, 93.4% owned a computer, and 82% owned cell phones. Women were 9% less likely to prefer online courses/high levels of learning technology:

Figure 15: Kvavit's Learning Technology Survey (Excerpt)



The United Nations predicted that 41% of the world's population would be carrying mobile phones by December 2009 (Brandon Hall 2009). This figure is no doubt larger now. Mobile learning has enjoyed rapid expansion in recent years and technological advancement has undoubtedly influenced this. Given the swift adoption of mobile technology in everyday life it is hardly surprising that it has 'spilled over' into learning environments (Calbraith & Dennick 2010). Although ownership of mobile devices does not necessarily mean that mobile learning automatically follows, it does mean that greater opportunity for it to happen exists. Kvavit (2004) states that most science-based disciplines prefer extensive technology in their courses, therefore it could be predicted that medicine/nursing will follow suit. This concurs with this thesis' findings in that a high level of animation and mixed interactive elements were desired.

Some educators recognise that mobile devices have not yet been fully exploited nor their limitations fully addressed (Cooper & Shufflebotham 1995). Others agree: "While the use of mobile devices has sparked the interest of an increasing number of researchers... our knowledge of learners' preferences for the mobile platform and their usage patterns remains limited" (Stockwell 2008, p270). This chapter will provide greater insight on these issues. However, I have noticed diverse and often dichotomous use of mobile learning in practice which serves to complicate emerging pedagogies.Thus educators encounter the same sort of pedagogical/evaluation difficulties with mobile LO learning as with desktop. This has left some educators wondering why their mobile learning is difficult to evaluate and asking 'Where did I go wrong?' (Felix 2005, Calbraith & Dennick 2009). Traxler (2009) recognises the difficulty in developing mobile evaluation strategies stating a need for more comprehensive, eclectic, and structured evaluation approaches based on sound/transparent principles. This chapter seeks to show exactly that.

8.2.2. Undertaking the Systematic Review

As previously stated, all mobile data was gathered and submitted to the same process as in 1B, thus a SR of generic mobile learning literature (together with specific ICU literature) was created and undertaken (see Appendix 42 for unique questions added to the data extraction sheet). Methodological rigor sheets were used as before (no modifications were necessary to ensure they included specific mobile aspects). Inclusion criteria incorporated 'any mobile learning *research*, evaluation or pedagogy'. It was deliberately wide to aid extraction of possible generic principles (and model formation). Consequently, material from any discipline/participant age/ location/type of course, and all types/years of all source data were eligible (providing that descriptions of advantages, disadvantages and effects of learning were reported).

An electronic search was made for all studies containing mobile learning research. This immediately showed a paucity of published literature so hand-searching was performed. 41 studies were found, and full text versions gathered when possible. Of these, 22 studies contained mobile research, but only 18 had the required inclusion criteria. Effectiveness of literature searching was checked and protocol rigor performed. This was judged 'adequate' taking into account the level of academic maturity present and the large narrative nature of the data. These 18 studies were forwarded for data extraction but only 15 studies had the required rigor level (N.B. 15 studies will not afford *robust* meta-analysis for ***all*** *mobile uses*, but following the SR protocol and using rigor sheets will afford *robust* meta-analysis for the use outlined in this thesis i.e. predicting effective approaches and will highlight any gaps in academic maturity and areas for further research).

Preliminary Analysis

There were 7 non-comparative studies, 3 comparative studies, 1 RCT, and 5 reviews found. These included 4 observational studies, 3 Controlled 'before-and-after' studies (1 contemporaneous, 1 case-controlled, 1 site-controlled), 3 qualitative studies, 2 cohort studies, 1 review, 1 focus group report, and 1 audit. Similar to 1B, baseline measurement and/or control group performance were not reported in many cases. Curricula type included 22 stand-alone packages, 2 stand-alone modules, 1 CMIC, and 1 internet-based course. 7 did not state curricula. Participants from nursing (3), medicine (2), computer science (2) mixed (4) were used (5 did not state discipline). Outcome measures included collaboration level; transition of passive to active learning; knowledge increase; user learning roles; and the benefits/ limitations of mobile technology.

The SR revealed several generic factors that influence effective mobile learning: Quick information, timely resource access at the point of need, changing educators/student roles, potential training delivery changes, constraints, broadening educational goals, and increased negotiated curricula. (Due to wordage constraints, full accounts of the descriptive results and statistics will not be detailed here, nor will 1C be written separately to 1B. Instead, each factor influencing effective mobile learning will be discussed below using 1B's format *together with* general/specific intensive care (ICU) mobile learning literature/research and a comment linking these to 1A findings where applicable (as in chapter 1C):

i) The importance of quick information - Brandon Hall (2009) describes the workforce as increasingly mobile thereby influencing learning delivery. Information can be viewed whilst waiting/travelling

thereby using otherwise 'dead' time. Although Smartphone examples are not available yet, PDAs have enhanced learning in the clinical environment by rapid acquisition of relevant information (White et al 2005). Rapid acquisition was very useful in the ICU setting where participants felt that quick procedure/drug checks were paramount for patients who can potentially deteriorate very suddenly;

ii) Timely resource access at the point of need is important. Learners are becoming accustomed to constant 'information connection' via mobile devices (Alexander 2004, Farrell & Rose 2008), allowing access anytime and almost anywhere (Holzinger et al 2005). They are versatile, customizable and portable (Baumgart 2005). Research has shown increased leadership skills and professional confidence when students' core knowledge and evidence-based information is reinforced in real-time when they need it (White et al 2005). However, few studies describe mobile learning within nursing practice and even fewer within nursing education (Farrell & Rose 2008). PDAs to date have provided resources at the patients' bedside, during clinic consultations and quick answers to practitioner queries. Specifically, Farrell and Rose (2008) describe 'e-tensive care units' where students used timely e-learning resources at the 'point of care'. The ICU nurses in this sample were excited at the prospect of having on-the-spot resources and immediately saw potential work benefits. They reported regularly queuing for computer use (accessible only when the patient's condition and nursing 'cover' allowed) thus considered mobile devices as 'greatly advantageous' (believing that timely access is not just helpful but crucial). This represents the first study to report smart/i/mobile Phone use. However, permanent adoption of mobile technologies would require several changes e.g. the educator role;

iii) Changing educators' roles - Since 2005 Nurse Educators have employed a variety of new learning technologies that increasingly focus on students (i.e. user-centred teaching). Consequently educators have taken a much greater role in course design and encouraging student information application in the workplace (Billings 2005). Educators' responsibilities, according to Pachler et al (2010), include helping students use new technologies effectively. Unsurprisingly, ICU educators were unaware of how their role may change but were readily able to highlight potential practical issues concerning 'confidentiality' and 'security of information'. Obviously, cultural changes are necessary in any context (Dearnley et al 2008), but encouragingly, Miller et al (2005) noted that established nurses (after having initial doubts) gave increasing support to students using mobile technology once they understood that they could access 'real-time' drug information. In this ICU sample, educators would need to adopt cultural change, teach learners how to use mobile devices, and access appropriate software (thereby adopting a much larger soft/hardware facilitating role);

iv) Changing student roles – Mobile devices allow students to "instantly construct their own learning for immediate application in real-world contexts" (Billings 2005, p343). Farrell & Rose (2008) undertook a pilot study to discover whether mobile devices would enhance students' pharmacological/clinical contextual knowledge, and to identify the effects of PDA use in clinical practice. Here the learner's role changed due to having to choose the most appropriate time/place and amount of times to access it. In short, it placed a greater emphasis on the degree of student choice and responsibility for learning. This ICU sample loved choosing their own LO subject and time of learning. They saw this as an enormous asset;

v) Potential training delivery changes - It has been suggested that mobile devices have the potential to change training delivery. However, handheld devices have been evaluated in a variety of clinical

environments since the 1990s (Farrell & Rose 2008) but have *not yet shown* a dramatic/widespread change to nursing/medical practice (De Ville, 2008). Neumeier's (2005) identifies one cause. When designing CAL-supported material, it became evident that a *systematic* investigation into the factors affecting mobile learning 'was missing and urgently needed' (hence this SR). In healthcare, real-time access to information at the bedside has the potential to 'improve care quality/safety, reduce adverse events and improving patient health outcomes' (Farrell and Rose 2008). However, when I reflected on most examples detailing training delivery changes, they entailed 'location' and did not involve people at all. People are at the centre of any change. In the ICU sample, nurses felt there was potential for training delivery changes providing that funding for devices was secured;

vi) Constraints - The adoption of mobile learning has often been constrained by slow networks and limited services. Hall (2009) describes 'anaemic devices', and the 'hesitancy of many organisations to purchase soon-to-be-obsolete hardware'. All of these constraints were found in this ICU sample, and cost/updating of devices was highlighted as a concern regarding ML becoming a long-term/permanent fixture. Buying mobile devices was not considered by the unit, however ICU nurses felt comfortable using their own mobile/smart phones providing they had been approved for use;

vii) Broadening educational goals - Felix (2005) believes broadening educational goals (i.e. lifelong learning concepts) has played a part in both the impetus and interest regarding ML adoption. Felix also believes that this inevitably leads to a 'social/cognitive constructivist learning paradigm', and highlights the pedagogical dilemmas present. Agostinho et al (2002) developed a tool to develop 'high-quality learning' based on generic constructivist learning designs but little has been reported about this. However, in practice I have witnessed the difficulties that development of pure constructivist approaches bring and have therefore sought to provide greater workable solutions. As previously stated, both deconstruction and linear development of LO pedagogy have met with limited results and conceptual barriers prevent further development. In ML it is therefore suggested that pedagogy should be developed *in conjunction* with the evaluation approach instead of being developed or appraised separately and then 'stuck' together. This helps 'iron out' any pedagogical problems early in development (Calbraith & Dennick 2009). In the ICU sample, broadening educational goals brought both advantages and disadvantages i.e. greater choice of relevant learning but also greater insecurity on whether learning choices were really appropriate, whether they would replace well-used traditional methods, and whether 'bought-in' learning would provide 'value for money';

viii) The increase of negotiated curricula - Kulusha-Hulme and Shield (2008) note that learning is no longer 'solely and carefully crafted by lecturers' due to learners being more mobile and motivated by personal learning needs.

During research technique (and LO) development for this thesis a great deal of consultation with stakeholders, educators, staff and students was undertaken regarding content to ensure high relevance. ICU nurses in this sample were consulted about whether they would use these LOs independently if given the opportunity. Many participants responded in the following ways: 'This is a great way to learn'; 'It makes you excited and *want* to know the answer'; 'It made me want to see the next bit of information'. When asked why they felt this way they explained it was because of i) how the LOs were constructed, ii) the use of immediate feedback guiding the user toward reasoned clinical decisions (motivating them to learn more), and iii) the ability to choose the most relevant aspects to

learn within any given subject. Those randomised to approaches containing scenarios liked the way scenarios developed further reasoning and application skills (Scenarios had been designed specifically to aid 'transferral of learned skills' to new contexts). Clearly this ICU sample correlates well with the factors that influence effective mobile learning however several mobile *disadvantages* were identified both in the literature and the sample:

i) Screen size: In the ICU sample this was evident. Some participants were concerned that night use may be difficult due to decreased lighting, and some thought that mobile text navigation requirements may be occasionally difficult. Back in 1995 researchers highlighted screen size difficulties affecting learners readily knowing where they were in documents used (Bartlette 1995). and the potential size of nested lists being limited (Cooper & Shufflebotham 1995). Waycott and Kukulsha-Hulme (2003) and Chehimi et al (2006) found several PDA limitations: small screen size for 'scan-reading', new/difficult text navigation requirements, flipping between reading documents and writing (no full keyboard therefore having to perform consecutive rather than simultaneous actions), and awkward note-taking. Despite considerable technological changes since 1995 (e.g. text input via i) virtual keyboard using 'letter-tapping'; ii) external keyboard connected via USB/IR/Bluetooth; iii) letter/word recognition translates into 'activated' letters; or iv) stroke recognition where predefined strokes represent various characters e.g. Palm's 'Graffiti'), screen size remained a small concern with ICU participants. Nevertheless they also felt that once they had got used to text navigation this would become easier, perhaps ceasing to be a problem at all. Several participants suggested that they would prefer on-screen 'site map indications'. Cooper & Shufflebotham (1995) suggest 3 ways to solve this problem – i) 'StretchText' (text grows/shrinks according to user preference/use); ii) text 'folding' (where more information is 'hidden', 'collapsed' or 'nested' underneath heading texts); and iii) screen rotation (to change page orientation e.g. iPad). Alternatively, bigger equipment could be used (e.g. iPad - icons grow bigger on screen to show the selected 'app' and employ 'pinch' techniques to navigate quicker). Finally, as Phase 2 showed 'site map indications' to increase learner satisfaction regarding learning preferences, simply adding 'page 2 of 4' (for example) may suffice;
ii) Ignoring the environment: Goth et al (2006) report mobile device users 'ignoring' the environment. Kristofferson & Ljungberg (1999) believe 'focus' and 'attention' can be potentially problematic in mobile learning. Ignoring the environment in ICU could be catastrophic for patient care. However, at no time did the focus of the ICU nurse get 'stuck' with the learning device. Each buzzer/change in patient status was picked up and acted on immediately. Several explanations are possible:

   a) Despite 'Top level' permission having been given for ICU research mobile learning was yet an 'accepted norm' for practice thereby users may have regarded it as 'having to slip it in where possible';
   b) 'not ignoring the environment' may also be explained by an unspoken discomfort about using mobile devices in the ICU context. In another context, Wishart (2008, p358) noted that language students felt that they "could not disrupt the established practice with the novel technology" due to the socio-cultural environment of the placements. If this is true of ICU learners, their attitude and behaviour may change if mobile device use was part of a permanent accepted practice. Stockwell (2008) highlights effects on establishments stating that workplace

support is paramount for successful use;

c) unit training encouraged a particularly good model of ICU nursing practice – i.e. finely-tuned recognition and response to patient needs; or finally

d) Goth et al's findings may simply not be generalisable to this context;

iii) Reticence to explore: In the ICU sample there was a reticence to explore mobile device capabilities beyond the research purposes. Wishart (2008) acknowledges that language teacher trainees did not fully explore mobile device potential because they 'were not yet confident in their pedagogical identities'. This hints that if 'standard pedagogies' were used user confidence may increase;

iv) Power short-comings: Chehimi et al (2006) found that mobile devices are limited to primitive battery power. Some PDAs lose stored content when batteries run down (requiring application re-installation e.g. Toshiba Pocket PC, Wishart 2008). Wishart therefore suggests one hour maximum usage when being deployed in wireless environments, thus limiting continuous use (Ganger & Jackson 2003). Since 2008 battery life has improved. More reliable devices may now get around this problem. Alternatively, use could be limited to 5-30 minutes at a time, or batteries recharged constantly using recharging 'cradles' (like ICU practice for other well-used devices);

v) Bandwidth short-comings: Cooper & Shufflebotham (1995) believe bandwidth may limit information retrieval speed. Kukulska-Hulme & Pettit (2007a) state that WiFi connection can be difficult/ impossible. Whilst acknowledged to be true, the ICU sample did not have a problem with speed/access using laptops but some did when using smartphones (N.B. In Miller et al's 2005 study 'speed of information access' increased student satisfaction):

vi) Memory capacity: Chehimi et al (2006) stated that mobiles are limited to 'diminutive processing power' and 'parameterised memory'. In this sample, this was not a problem (and devices can always be plugged into other equipment allowing 'higher specification peripherals' if the learning content requires greater power - e.g. using serial ports and/or USB cables, Cooper & Shufflebotham 1995. Conversely, an extension card with an Ethernet port and/or RJ-45 adaptor can be used, Wikipedia 2010). Extending desktop function with pen drives has become normal practice - the only difference is that mobile devices with plug-ins are not yet considered 'everyday equipment' for most. It is therefore unfamiliar;

vii)Security issues: There is a potential for security/privacy issues to arise between users (i.e. confidential patient information storage or if using wireless exchange of data transmitted between client/WAP). This can be overcome by using a secure/private connection wireless link (Ganger and Jackson 2003). Although these potential problems were not actual ICU problems it is noted that other material may have changed this. Teaching staff quickly identified that policies would have to adequately incorporate new technology to prevent breaches in care, cheating during on-line staff tests/attainment of competencies. Wilkinson et al (2006) advocate the collection of all computer IP addresses in the required setting for the required function (i.e. test/exam) and combining these with a system where the server is able to reject log-in requests from unauthorised persons and prevent re-entry to an exam paper once the page is 'exited'.

There was one ICU concern not found in the generic literature. Some nurses were initially concerned whether mobile devices could potentially interfere with pumps, monitors and equipment. I explained that equipment compatibility was paramount for patient safety, that advice had been sought from a unit technician before research commenced, and that safety would be assured for this type of learning if

implemented permanently. For healthcare settings I suggest a formal risk assessment be completed each time a different use or make of mobile device is required. As the disadvantages found did not necessarily preclude ICU mobile use, LO learning was deemed fit-for-purpose providing the above recommendations are considered. Several *advantages* of mobile learning were also noted:

i) Waycott (2002) found viewing applications with fingers/stylus was quick/easy for users. Even ICU nurses that had not previously used Smartphones felt that choosing 'apps'/web-links this way is relatively easy (providing they had training);

ii) Some devices can communicate with other local PDAs/computers by 'beaming' or 'synchronising' data (Cooper & Shufflebotham 1995). The ICU nurses were unknowingly familiar with this method in another context (BARS blood label system). If devices use Bluetooth technology, Bluetooth compatibility should be checked/maintained or it may become a major problem;

iii) When used in language teaching, Nah, White and Sussex (2008) found that smartphones enhanced listening skills and encouraged students to actively engage in learning. ICU nurses were actively engaged. It is suggested that that any method promoting active engagement has merits;

iv) Waycott (2002) found that the portability of mobile devices a great asset. Indeed, these devices "are constantly available to their users due to them being 'personal' and constantly to hand" (Cooper & Shufflebotham 1995, p2). The ICU sample used mobile/i/smartphones and laptops and felt that the portability allowed use at the bedside, outside work, and whilst travelling;

v) The main navigation method is for mobile learners to select the links they are interested in. Kukulska-Hulme & Pettit (2007) describe the main advantage of Smartphones as being 'self-service' education i.e. led by students' own learning needs with tutor support available. Cooper & Shufflebotham (1995) believe this is particularly true if devices are set up to access learning material or perform certain actions without the need for 'filling in' information first. All ICU nurses found the simple LO choice method easy to use;

vi) Brandon Hall (2009) suggests that most learners already own mobile phones so there is already a psychological advantage for them regarding learning 'ownership' and 'direction'. In ICU, nurses expressed great enthusiasm however further research is required to establish whether greater learning ownership levels are gained using 'self-owned' versus 'loaned' devices;

vii) Chehimi et al (2006) talk about 3D applications for Smartphones. It is easy to envisage how 3D graphics could aid ICU learning, particularly for anatomy/physiology. This would have undoubtedly enhanced the LO images used in the ICU sample;

viii) Wishart (2006) says that all handheld recording methods are popular, and students particularly value the ability to capture 'on-the-spot'events/reflections through video-recording. This facility was not used in the ICU example permitting no direct comparison. However, consented recording of patient assessments could be used as evidence of competencies gained. Never-the-less, it is suggested that ethical consideration of patients' feelings/dignity should be considered at length if this use is desired. Many patients/relatives may find this type of 'care' a little too intrusive.

In conclusion, 1A and 1B above were well integrated and displayed many mobile advantages. Mobile findings proved to hold similar levels of complementarity; enhanced insights etc as desktop learning thus this method can be used for both desktop *and* mobile LO learning. Indeed, many of the potential problems listed here did not become problems in practice when researched. However, many mobile advantages have not yet been *fully exploited* either. They clearly have the potential to dramatically

change the way both formal and informal learning is supported (Waycott 2002). *Recommendations* - When creating pedagogy/evaluation for mobile learning, issues outlined above should be considered. There should not be major problems when creating these LOs *providing that* i) devices, software, text size and content are all carefully considered and specifically designed for mobile use; ii) content is chosen according to the specific intended use; iii) the devices are installed, monitored and serviced according to health and safety regulations); and in healthcare settings iv) patients/relatives have given informed consent where appropriate.

8.2.3. 1B Results/Directions for Phase 2 Research (According to SCIE SRES1, My System and Statistical Analysis)

In Chapter 5 (Project 1B) results and directions for Phase 2 testing were presented separately. Due to lack of wordage participant, study and intervention statistics will not be presented here (but are available on request). Study design type and descriptive statistics will be discussed and mobile SR results will be presented together with recommendations/directions for further testing to be performed later in this section (N.B. Most mobile learning studies achieved level 1 Kirkpatrick impact. However, as Kirkpatrick was shown earlier not to measure rigor and as identification of top/bottom-performing mobile approaches are required, Kirkpatrick impact will not be discussed here).

Study Design Type - LO Research

In Table 37, 46.66% of studies were non-comparative, 33.33% comparative, 13.33% qualitative, 6.66% reviews. Case-controlled CBA came top (13.3%) followed by ITS (6.6%). (Red shows top results, yellow shows systems that came second, and blue shows those that came bottom).

Table 37: Study Design Type - LO Research

Ref	Type	Details	Overall evidence level (My system)	Overall evidence level SCIE SRES1
M001a	B - comparative	Cohort	39.99% at type III, rigor level 1, 47.82% at type IV	High – 77.07%
M001b	A - non-comparative	Observation	47.82% at type IV	High – 77.07%
M002	D - other		9.09% 2/22 at type IV - rigor level 1	Low – 16.96%
M003	A - non-comparative	Post-test	49.99% 10.5/21 at type IV	Med – 49.39%
M004	B - comparative	ITS	53.84% 7/13 at type III	Low – 32.58%
M005	B - comparative	CBA – case controlled	67.85% 9.5/14 at type III	Med - 66.29%
M006	B - comparative	CBA – site controlled	38.46% at type III	Med – 34.06%
M007	D - other		25% 3/12	Med – 60.68%
M008	A - non-comparative	Observation	61.9% at type IV	Med – 63.83%
M009	A - non-comparative	Focus	36.36% at type IV	Med – 55.35%
M010	C - review		64.28% 4.5/7 at type IV 14.28% type III	Med – 44.93%
M011	B - comparative	Cohort?	45.83% 5.5/12 at type III	Med – 66.21%
M012	A - non-comparative		39.99% 6/15 at type IV	Med – 36.68%
M013	A - non-comparative	Observation	53.84% 7/13 at type IV	Low – 32.18%
M014	B - Comparative or A - non-comparative?	Audit	17.39% 4/23 at type IV	Med – 50.44%
M015	B - comparative	Case-controlled CBA	86.66% at type III	High – 85.49%

As in 1B, statistical analysis was performed on mobile data using excel's 'NORMINV' function for each level (10%, 25%, 50%, 75%, 90% and 95% - Table 38):

Table 38: NORMINV Comparisons for Study Design Type

LEVEL	SCIE	MY SYSTEM
10%	34.81274	22.2594568
25%	45.38806	32.91480321
50%	57.138	44.75366667
75%	68.88794	56.59253012
90%	79.46326	67.24787653
95%	85.79216	73.62467246

As I had performed a SR of all existing literature, and the sum of the scores were taken, it is assumed that the distribution for the whole *known* population was estimated. The normal distribution curve was plotted and 10%, 25%, 50%, 75%, 90%, and 95% for both my system and SCIE were added. However, NORMDIST was used to give a normal distribution (for the mean and SD) to complete the calculation. This showed 0.34 (SCIE), and 0.61 (my system).
The maximum rigor confidence interval (p=0.05, 95% level) is 85.79 (SCIE) and 73.62% (my system) therefore anything in Table 38 over these values is both performing well and statistically significant at the 95% level (as shown by Table 39).

This means that the best study according to my system was M015 CBA performing above the 95% level (if measured at Level III). According to SCIE it is M015 at the 95% level, followed by M001 just under the 90% level. Although most percentages in Table 39 were different between my own and SCIE's system, the placing was nevertheless similar - they both identify identical top/bottom-performing approaches. Placings in the middle differed in relation to how comparative the study was. This may suggest that Case-controlled CBA perform well when used for mobile LO learning but bottom-performing approach comparison is required before assumptions can be made.

To test the null hypothesis, bottom-performing studies were also included – i.e. any studies performing under the minimum (10%) confidence level of 34.81 (SCIE), or 22.25 (my system). This showed M002 (qualitative study) as the worst in my system and M015a (Case-controlled CBA) scored only marginally higher. As CBAs also come top this suggests once again that it is the *way* the research is *conducted* that counts and that no particular research method is best for mobile learning. Changing the delivery format has therefore made no difference to Chapter 1B findings thus (for the sake of parity and ease of comparison) all mobile approaches to be tested in Phase 2 will be performed using observation. To summarise, the results of the SR meta-analyses were as follows in Table 39:

Table 39: Comparison of Systems - LO Research

Placing	SCIE placing (%)	Statistical SCIE placing (NORMINV)	My placing (%)	Statistical my placing (NORMINV)
top	M015b - Case controlled CBA	M015b - Case controlled CBA	M015b - Case controlled CBA	M015b - Case controlled CBA
2nd	M001 - Cohort	M001 - Cohort	M005 - CBA	M005 - CBA
3rd	M005 - CBA	M005 - CBA	M004 - ITS	M004 - ITS
4th	M011 - cohort	M010 - review	M011 - Cohort	M011 - Cohort
5th	M008 - observation		M001 - Cohort	M001 - Cohort
2nd to last	M013 - observation	M013 - observation	M015a - Case controlled CBA	M015a - Case controlled CBA
Last	M002 – qualitative observation	M002 – qualitative observation	M002 – qualitative observation	M002 – qualitative observation

LO Pedagogy

Table 40 shows M015b (McClure et al's Constructivist approach) as 'top-performing' (95% level). M008 (Danesh and Prinsen) and M010 (Dix & Jones' deconstruction) only perform well at the 75% level, thus McClure's approach will be tested in Phase 2. M002 (Piaget 1936 & Dewey 1938) was bottom-performing (below 10% level) therefore Piaget's experiential learning will be tested in Phase 2.

Table 40: Comparison of Systems – Pedagogy

Study ref	Placing	Author/approach used	Approach
M001		not stated	collaborative
M002	Last	Piaget 1936 & Dewey 1938	experiential
M003		not stated	not stated
M004		not stated	not stated
M005	2nd	Not stated	Not stated
M006		not stated	Experiential?
M007		not stated	not stated
M008		Danesh et al 2001, Prinsen et al 2007	collaborative
M009		Naismith et al 2004	collaborative
M010	3rd	Dix 2003; and Jones 2006 – information shaping the users world	deconstruction
M011		Engestrom 1996 – activity triangle model	Vygotsky 1978 & Leont'ev 1979 Activity theory
M012		Not stated	not stated
M013		not stated	not stated
M014		not stated	not stated
M015a	Second to last		Not stated
M015b	top	constructivist - McClure & Gatlin 2007, Taylor Maor 2001, Taylor, Dawson & Fraser 1995, ?, Chang & Fisher 2001	constructivism

LO Evaluation

In Table 41, the top-performing study was M015b therefore 'Emancipation/reflective thinking/co-participation' will be taken forward for Phase 2 testing. Details for M005 and M010 were not given by the authors. The bottom-performing study was the 'just in time' experiential learning model. This too will be taken forward for Phase 2 testing.

Table 41: Comparison of Systems – Evaluation

Study ref.	Placing	Author/approach	Approach
M001a and b		not stated	collaborative
M002	Bottom	just in time learning	experiential
M003		not stated	not stated
M004		not stated	not stated
M005	2nd	not stated	unknown
M006		not stated	experiential
M007		not stated	Not stated
M008		not stated	collaborative
M009		not stated	collaborative
M010	3rd	not stated	Not stated
M011		not stated	Vygotsky 1978 & Leont'ev 1979: Activity theory
M012		not stated	Not stated
M013		not stated	Not stated
M014		Own - integrated	constructivism
M015a	Second to bottom	not stated	constructivism
M015b	top	not stated	Emancipations, reflective thinking, co-participation

8.2.4. Directions for Phase 2 (According to SCIE SRES 2)

Table 42 shows that there were 17 studies - 4 high, 11 medium and 3 low (0-33%=Low, 34-66%=Medium; 67-100%=High). Top (red), Second (yellow), Bottom (Blue).

Table 42: Directions for Phase 2 (According to SCIE SRES 2)

Study Ref.	Type of study	A Basics complete	B Appropriateness of sample design, data collection/ analysis	C Relevance of topic focus to review	D Quality	E Overall weight of evidence	Type of evidence (taken from SR rigor sheets)
		SCIE SRES2					MY SYSTEM
M001a	B - cohort	High - 83.3%	Med - 50%	High – 100%	High – 75%	High – 77.07%	39.99% 6/15 at type III = rigor level 1
M001b	A – observation	High - 83.3%	Med - 50%	High – 100%	High – 75%	High – 77.07%	42.87% 11/23 at type IV
M002	D - other	Low – 0%	Low – 0%	Med – 42.85%	Low – 25%	Low – 16.96%	9.09% 2/22 Type IV - rigor level 1
M003	A – post-test	High - 83.3%	Low – 25%	Med – 64.28%	Low – 25%	Med – 49.39%	49.99% 10.5/21 at type IV
M004	B - ITS	High - 74.99%	Low – 12.5%	Med – 42.85%	Low – 0%	Low – 32.58%	53.84% 7/13 at type III
M005	B - CBA	High – 100%	Med – 43.75%	High 71.42%	Med – 50%	Med - 66.29%	67.85% 9.5/14 at type III
M006	B – CBA site controlled	Med - 66.6%	Low – 12.5%	Med - 57.14%	Low – 0%	Med – 34.06%	38.46% at Type III
M007	D – other	High – 100% 2/2	Low – 33.3%	Med – 42.85%	Med – 66.6%	Med – 60.68%	25% 3/12
M008	A - observe	High – 100%	Med – 42.85%	Med – 50%	Med – 62.5%	Med – 63.83%	61.9% at type IV
M009	A – Focus?	High – 100%	Med – 42.85%	Low – 28.57%	Med – 50%	Med – 55.35%	36.36% at type IV
M010	C – r/v?	Med – 50%	Med – 58.33%	High – 71.42%	Low – 0%	Med – 44.93%	64.28% 4.5/7 at type IV Or 14.28% 1/7 type III, 9.09% type I
M011	B – Coho?	High – 91.66%	Med – 37.5%	High – 85.71%	ɐd – 50%	Med – 66.21%	45.83% 5.5/12 at type III
M012	A – observe	Med – 58.33%	Low – 31.25%	Med – 57.14%	w – 0%	Med – 36.68%	39.99% 6/15 at type IV
M013	A - observe	High – 74.99%	Low – 0%	Low - 28.75%	w – 25%	Low – 32.18%	53.84% 7/13 at Type IV
M014	A – Audit?	High - 100%	Low – 12.5%	High – 89.28%	w – 0%	Med – 50.44%	17.39% 4/23 at Type IV
M015a & b	B – CBA case controlled	High – 100%	Med – 56.25%	High - 85.71%	gh – 100%	High – 85.49%	86.66% at type III

M015a and b came out top according to SCIE's SRES 2, followed by M001a and b (High) and M0011 (Medium/high). M002 was last, and M013 and M004 were joint second-to-last. According to my system (methodological rigor sheets, M015 came top followed by M005 (High) and M010 (medium/high), M002 came bottom, M014 second to bottom. When comparing SRES2 results with SRES1 and my system they almost concurred identically. Were they did not concur (middle scores) *the level of robust argument* appeared to be the causative factor e.g. *Evaluation:* M003 scored HIGH, M004 MEDIUM/HIGH; and *Pedagogy:* Leont'ev (1979) and Vygotsky's (1978) Activity Theory scored MEDIUM regarding robust level of argument measures. This thereby changes their overall SRES2 position. As SRES2 concurs with SRES1/my system findings, the approaches already identified for Phase 2 testing in Section 8.2.3 will therefore be taken forward.

8.2.5. Discussion

It can be said that this approach (usability study, grounded theory, systematic review, grounded theory development) when used with SRES1/my rigor sheets is capable of finding and robustly evaluating all existing mobile pedagogies/evaluation approaches. Despite Chapter 1B pedagogies working well with mobile learning (see 1A previously in this chapter), different top-performing pedagogy/evaluation approaches are found for mobile learning when the *whole process was completed (1B, Phase 2)*. This could be accounted for as follows - Pedagogies may be *interchangeable* between mobile and desktop delivery. It is possible that the mobile pedagogies found may work equally well with desktop delivery but this has not yet been fully researched. If this is the case it can be said that effective mobile or desktop pedagogies/evaluation approaches can be used for both delivery formats (i.e. approaches are interchangeable). If mobile pedagogies/evaluation approaches are found to be ineffective/non-appropriate when using desktop delivery they are therefore *not* interchangeable thus cannot be used for desktop delivery. Further research is required.

Further format testing has never-the-less shown that there is no need to change the way top and bottom-performing approaches are *identified* when delivery format is changed to mobile learning as it appears to be just as reliable when used for desktop access. It can also be tentatively said (tentatively because of the small number) that changing format does not necessarily mean having to change the pedagogy/evaluation providing that i) comparable research has taken place regarding the specific approaches desired (to ensure that change in format does not introduce extraneous variables; ii) no substantial differences are noted between delivery formats (i.e. in this case desktop and mobile formats). As the change in format did not adversely affect the research and concurred with desktop grounded theory very closely, mobile learning complied with this first proviso. Differences found were minor but nevertheless require discussion to estimate whether the second proviso was fulfilled. These differences between desktop and mobile LO learning related to activation of prior learning and transferral of knowledge/skills.

Activation of Prior Learning

This appeared to be more important in mobile LO learning than desktop and appeared to be linked to the size of the page (i.e. smaller screen size equating to lower learner confidence due to learners less able to keep track of where they are in the learning. Activation of knowledge becomes more important as it helps the learner focus whilst navigating thereby helping learners to link information). Activation of prior knowledge is advocated by Ausubel (1968), Keppell et al 2002) and Schmidt (1993) as new knowledge can be 'subsumed into existing concepts/bodies of knowledge'. Keppell et al 2002) believe that this creates 'learner readiness' by asking learners to generate hypotheses. Schmidt (1993, p424) goes further indicating that the *extent* of activation determines the type of new information and how much of it is processed. In this thesis activation of prior learning was not actively planned for during desktop/mobile research but witnessed indirectly during usability observation/interview. In both cases, top-performing pedagogical approaches had a greater degree of 'knowledge/expectancy activation' than the lower-performing approaches, evidenced by the way learners spoke about the learning e.g. 'I think I know what the next bit is going to be.... I did this a couple of months ago... ah yes... it is!'

Again, this process was more important in the mobile (versus) desktop LOs. Mayer (2001, p161) talks about the 'Individual differences principle' where multimedia "design effects are stronger for low-knowledge learners than for high-knowledge learners (in retention and transfer tests), and for high-spatial learners than low-spatial learners (in transfer tests only)". He believes that high-knowledge learners compensate for lack of information by activating prior knowledge. He also states that "high-spatial learners have the capacity to integrate verbal and visual representations whereas low special learners devote their cognitive capacity to holding the presented images in memory" (p161). If this is the case, it would seem logical that these design effects are inversely related to screen size - the smaller the screen the more difficult it is to hold all elements in the working memory. This gives further weight to the importance of prior learning activation outlined above.

Schmidt (1993) believes that storage/retrieval of information is enhanced when material is 'elaborated upon'. It is worth noting that material cannot readily/quickly be elaborated upon if using mobile learning as 'distance' learning. Jurczyk et al (2004) believes three important instruction dynamics are changed: i) the modalities of communication (less modalities means that clarity is paramount, and working relationships take longer to develop Harasim, 1988; Sproull & Kiesler, 1992), ii) the management of time (communication becomes 'asynchronous' therefore regular checking up on students particularly those with no experience of distance-learning is required to ensure that they understand what is expected from them); and iii) the formats for assessment. Online learning should not be limited to remembering facts but should be designed to show application of skills/knowledge. It is therefore suggested that Educators can use IHEP standards to help plan distance learning modules. The notion of material being 'elaborated upon' was particularly noticeable in the LO using Chapter 5 (1B) pedagogies that included the scenario because students had to develop their own clinical principles to explain/deal with medical emergencies/deteriorating patients (I saw the ability to transfer this knowledge of the principle to another context as crucial for both nursing and medicine because applying theory to real-life practice is essential). Gagne (1986) describes this process as 'fine-tuning or changing schemata', Ausubel (1960) as 'Intellectual scaffolding', and others as cognitive 'chunking'. In short, this research enabled learners to place new information in the best place for them to link theory and practice. Clinical reasoning developed as a result (The LOs will not let you continue until the learner has given reasoned and correct principles). This was evidenced by learners reaching the end of the scenario and by the decreasing number of times certain questions were attempted before a right answer was achieved. The learning designs were seen to have aided better learning. This is not surprising as Anderson (2003, p 21) describes learning designs as "the instructional 'glue' that holds various parts of an instructional episode into a complete learning experience". Thus it is recommended that LOs using this format have a structure that supports student development and supports student 'elaboration on material'.

When searching the literature for possible concepts to support these deductive evaluations, a model based on 'orientation to study' by Schmeck, Ribich, and Ramanaiah (1977) was useful. This model combines reflective and agentic processing. The result is distinct transferability with memory durability and fact retention. This suggests that when LO formats incorporate 'elaboration of material' features, they are likely to induce 'longer lasting' learning due to the reflective processes involved. In summary,

as no *substantial* changes are needed it can now be said that changing format does not *necessarily* mean having to change the pedagogy/evaluation for ICU nursing. Desktop pedagogies/evaluation approaches can be used providing that mobile format differences are taken into account.

8.3. Phase 2: Testing of Pedagogy and Evaluation Approaches

8.3.1. Grounded Theory

The top-performing mobile approaches from Phase 1A were taken together with bottom-performing approaches and were each given the same null hypothesis as before. 35 ICU nurses were randomly selected from 110 identified to participate in the study, 31 agreed covering 36 LOs. Each participant was blinded/randomised to one of the 4 approaches. Romero & Wareham (2009, p4) ask question we all should be asking "What type of learning do we want from mobile technologies?" They compare 'permanent behavioural change' versus 'speedy problem-solving'. For this sample (ICU) it is suggested that *both* are required: speedy knowledge to deal with immediate problems and this new information should then become part of practitioners' established practice. When comparing this phase with 1A results, lesser degrees of learning satisfaction was noted on the bottom-performing approaches, and knowledge score ratings were 30-40% higher (on average) when using the top-performing approaches. The results are also very similar with the exception that 9.6% of this sample felt that *extra* links would enhance the learning further. Phase 2 results are as follows:

Findings: General

General trends included the following:

- 70% of participants preferred information presented as images/animations whilst the remaining 30% preferred written/text elements (summarised information, simple terminology/ definitions);
- 97% of participants liked 2 or more elements per screen to help focus learning and prevent boredom
- 50% of participants preferred the inclusion of test elements, the remaining 50% appeared to favour any method that aided the practical application of learning;
- Minor changes wanted in elements were equally distributed between visual (i.e. larger pictures) and written elements (i.e. less/more text, more labelling);
- Most participants felt navigation through the learning material on-screen was good No negative comments were noted.
- Comments about content of diagrams were all positive - a small percentage of participants (3%) wanted slightly larger images.

Findings: Underlying factors

As coding progressed it became obvious that the same 8 main core codes (with regard to underlying pedagogical and evaluation factors) found in Phase 2 Section 7.3 were found here also. The descriptions of the categories/codes are a little different but this is not surprising as these are different participants with different delivery format. However, many similar ideas appear:

Information overload – When elements have the right speed and structure active learning is increased and overload is decreased due to learners feeling that they have had the time to take in the learning.
Time – The structure of pedagogy and evaluation used decreases overload and increases the speed of understanding/clinical reasoning.
Monotony – Images and test elements (i.e. good pedagogy) increases level of active learning.
Interest – images increase the level of active reflection upon learning.
Attention – Increased interaction increases level of active learning.
Control – Clear information and user-friendly navigation helps the learner decide where they want to go and what they need to learn next (i.e. gives pedagogical direction).
Application of learning – 'Good' pedagogy increases understanding which increases reflection which increases motivation. 'Good' evaluation increases application of learning due t the structure and feedback mechanisms.
Motivation – Good text gives the participant 'something to aim for' (i.e. learning goals).
Participant learning preferences – A good mix of elements shows participants how and what is wrong in their understanding. This increases the level of active learning, gives direction as to what learning is needed next which increases the application of learning which increases good linkage of knowledge.

Findings: Main Factors

Good pedagogy

- Increases understanding
- Helps apply theory to practice
- Can be applied to any learner level

Good evaluation

- Develops clinical reasoning
- Helps learners understand what they know

When comparing these main factors with Phase 2 section 7.3 findings similar themes are evident – application of learning, all levels of learner etc despite the different language used to describe them. This shows good parity to and within the two nursing samples ('Phase 2 section 7.3' and 'further format testing phase 2'). Categories/answers were checked with participants who verified that these were exact reflections of their experience. Data collected was 'theoretically sensitive' and gave accurate meaning to and categorisation of the data. Again, tight linkage between categories was seen. In short, mobile findings verified the desktop grounded theory (GT).

8.3.2. Theory Generation

Again, comparing associations within and between single items that relate to the same category and all relate separately to the core index - e.g. pedagogy (Table 43) and evaluation (Table 44) were formed:

Table 43: Comparing Associations Within/Between Single Items - Pedagogy - Phase 2 Theory Generation

Concept= pedagogy	Minimum = poor pedagogy	Maximum = good pedagogy
Quality of mixed elements	When text and image elements are of poor quality the speed of reading is impaired by the feeling of being distracted and overloaded	When text and images are of high quality clinical reasoning develops quicker which increases enjoyment, motivation, and reflection. This in turn increases understanding.
Clarity of explanation	Understanding decreases when explanation is unclear	Good pedagogy explains why something works therefore increases understanding by making clinically relevant points clear and making it 'real' for the learner
Clarity of practice /theory links	Does not make practice theory links clear	Applies things to real-life providing application and transferral of knowledge
Learning levels	Aimed at one level of learning thus may totally 'miss the target' for some learners	Caters for all levels of student

Table 44: Comparing Associations Within/Between Single Items - Evaluation - Phase 2 Theory Generation

Concept = evaluation	Minimum = poor evaluation	Maximum = good evaluation
Assessment	No assessment is possible	Assesses the learner
Clinical reasoning level	Learner does not progress clinically	Develops clinical reasoning due to increased motivation when feedback language is good
Learner level	Learner level cannot be ascertained	Frequent evaluation ascertains learner level and saves tutor time. All levels of learner are found due to the evaluation's structure
Learning objectives	Learners feel unable to judge their progression	When learning objectives are clear this gives learners the ability to self-evaluate

Although expressed differently, 'element mix', 'application of learning', 'feelings of overload', and 'catering for all levels of learners' are common to both this sample and 1A mobile findings. Good pedagogy (giving the learner the ability to self-evaluate) is an identical theme. In both cases the structure of evaluation appears crucial. Clusters of relationships *between* different codes were also considered: all theoretically relevant relationships among the concepts were extricated for each core index (pedagogy and evaluation) and their underlying factors (Table 45). Feelings of overload and personal learning preferences were linked to increased understanding. Timing of pedagogical/ evaluation elements appear crucial and links between images interest, monotony and active learning were noted. Learning choices feature more highly in this sample than desktop. Specific examples of how components affect pedagogy/evaluation are identified.

Table 45: Clusters of Relationships Between Different Codes - Phase 2 Theory Generation

Concept	Maximum=appropriate level	Minimum=not appropriate level
Information	The right structure and timing of elements increase active learning by allowing learners to 'digest' the learning	Feelings of overload are increased
Time	Good structure and timing of text, pedagogy and evaluation increase the speed of understanding and clinical reasoning	Feelings of overload are increased
Monotony	Good pedagogy and evaluation (i.e. good quality images and test elements) keep attention and increase active learning	Learning is not active and lacks reflection
Interest	Good audio/images provoke active learning and increase reflection	Learning is monotonous
Attention	Right level of images/ interaction increases attention therefore increasing active learning	Learning is monotonous
Control	Clear information/user-friendly navigation helps the learner decide where they want to go and what they need to learn next	Unclear information leads to less control over learning
Application of learning	Good pedagogy increases understanding by increasing reflection and motivation. Good evaluation increases the application of learning due to good feedback structure	Lack of evaluation/summarised information decreases application of learning
Motivation	Good text gives learners something to aim for (learner targets)	Bad text: learners are distracted, speed of reading is impaired and their desire for feedback increases
Participant learning preferences	Good mix of elements shows learners what/how something is wrong. This increases active learning and gives direction to the ensuing learning needs increasing the application of learning and linkage of knowledge.	Bad mix increases confusion and lack of application.

8.3.3. Hypotheses

Hypotheses formed by this mobile/ICU group were:

- A good mix of LO elements result in learning errors being identified, increased awareness of learning direction and needs which in turn allows a greater application of learning, linkage of knowledge, and development of clinical reasoning;
- Unclear information/Layout due to lack of clarity or information decreases active learning resulting in learners feeling that they do not have the time to 'digest' the learning;
- Clear learning objectives give learners the ability to self-evaluate;
- The right level, structure and timing of elements increases active learning by allowing the learner to digest the learning resulting in increased speed of reading, understanding, clinical reasoning, which does not distract or overload the participant.

Phase 2 findings not only fitted into 1A mobile theory but gave it even greater explanatory power by adding detail. It is clear that both sets of hypotheses are mostly complementary therefore generic principles may work for both desktop and mobile delivery format (despite using different approaches found using SRs).

8.3.4. Hypothesis Testing

Initially, the GT produced in Chapter 1A was considered incomplete until the SR had been conducted and findings compared. This theory has now been tested using seven different populations (Medicine x 2, Nursing x 3, IT, Science, Laypersons, Education, Science Education) in 3 locations (1 x Midlands and 2 x Eastern areas) and now 2 delivery formats. Findings were very similar (and had innate transferability of this method across topics, disciplines, and locations).

Minimising and maximising the concepts helped to form not only GT boundaries and judge rival explanations but over the course of the thesis illustrated the convergence of a construct. However, it would be premature to claim that top-performing approaches can be used for *all* LO learning whatever the delivery format. Projects with large numbers and similar hypotheses were sought but not found, however some confirming instances and their conditions were: i) It has been well-documented that multi-modal learning within e-learning increases knowledge retention, interaction and participant interest which appears to support the 'Mixed elements' finding; ii) No literature has reported 'section links not being detected due to learning preferences' in mobile learning, however the increased need for page positions has been discussed; iii) Although few papers comment specifically on mobile learning objectives, it is well documented that good, appropriate and measurable learning objectives result in effective learning; iv) 'Good labels do not overload' – No specific nursing examples were found, however intuitive mobile designs appear to provide the most effective usability.

In summary, *full* hypothesis development was not possible with the ICU example as there were few specific mobile learning instances recorded with which to provide further confirming/disproving instances/conditions. However, given that most GT stages were possible (and that nothing else presently exists) this implies that there is sufficient grounds to examine inherent generic principles and build a tentative LO learning model because: i) similar codes, grounded theory and ideas about good

pedagogy/evaluation are found in both mobile and desktop delivery regarding LO learning despite being blinded; ii) no input was given by me - theory was generated directly from verbatim comments; and iii) the rigor/reliability of the mobile SR findings have obvious parallels to desktop SR findings. Excitingly, given that each group tested in this thesis were i) given completely 'free-reign' during observation regarding comments on the learning and ii) free choice of how they used them during Phase 1, it is likely that generic principles do exist. This evidence is very promising and provides a great basis for further development.

8.4. Formation of Generic Principles

An emphasis on 'good' pedagogy/evaluation was pursued within the further format testing Phase 2 to see whether generic principles could be formed for mobile learning. Participants thought good pedagogy i) "makes you prioritise your care and use it like in real life", ii) "is clever because combines scientific knowledge and nursing practice", iii) "helps you know what you've learned", iv) "reminds you of what you've forgotten and helps you re-apply it", and v) "is interesting and keeps your attention active which is easier than learning from books". Participants thought good evaluation was when i) both coursework and summative assessment are used as "this is good balance for assessment", ii) when the evaluation includes questions as "this retains knowledge", iii) when the evaluation guides users towards reasoned answers "as the information given can be used practically", iv) when the evaluation has MCQs "as this requires reasoned answers which can be built upon", v) when the evaluation shows that new learning has built upon existing knowledge, vi) when the evaluation itself aids application of the new knowledge, and vii) when evaluation is set out using various different contexts/methods as "it focuses the attention on important parts of the learning". These created the following principles:

Effective mobile evaluation includes *(Generic principles):*

- both formative and summative assessment to achieve a good balance
- questions to aid knowledge retention
- a demonstration that new learning has built upon existing knowledge
- the easy application of new knowledge
- focused attention on the important parts of the learning

Effective mobile pedagogy *(Generic principles):*

- Reminds the user what they have learned and how to apply it
- Keeps attention active

Effective mobile pedagogy *(Principles unique to discipline – in this example, ICU Nursing):*

- Encourages the same prioritisation of care as in real-life
- Combines scientific knowledge and nursing practice

Effective mobile evaluation (*Principles unique to discipline – in this example, ICU Nursing):*

- Has question formats that guide of users towards reasoned answers

Thus this process was able to 'distil' several principles directly from GT findings. These were then checked against SR findings to ensure that each top-performing pedagogy/evaluation approach had these attributes (where differences existed these were not considered 'generic'). Anderson and McCormick (2005, p1) believe "The judgment of pedagogic quality has to be "principled", in that any particular decision to create and use e-learning should be underpinned by some agreed principles of good teaching and learning". They also state that these principles may be derived from "distilled practice as found in educational institutions" (2003, p20). Since my generic principles had a very wide SR practice 'backdrop' (all subjects in existence), it is suggested that these principles can be used as a design basis to build LO learning on *any subject* within the environment it is formed for (in this case mobile learning in ICU nursing), or the process replicated this process to gain generic principles for other populations/subjects/delivery formats. When comparing the most effective LOs found in this thesis with the work of 'deeper learning principle' authors (i.e. Carmean 2002, Weigel 2002, Dabbagh 2000 & 2003, and Shulman 2002), it is interesting to note that similar themes for learning are seen. These include contextual/authentic/active learning/knowledge; real-world problem-solving; and searching for underlying principles.

Anderson and McCormick (2005) state that there is an implicit assumption that the more principles the learning has the better the pedagogical quality. Therefore few principles mean that some important pedagogical principles may be missing. Anderson and McCormick (2005).recognize that e-learning developments may not embody all principles and therefore advocate learning activities to support and address the 'missing' principles. McGee (2003, p6) believes that requisites for the discovery of principles exist: "Although learning may be supported through instructional strategies it is difficult if not impossible to design for missing principles without systems that can provide individualized feedback and interaction that is personalized, responsive, and immediate". Feedback in my LOs was immediate and it was the same for everyone. Its structure, however, gave learners the impression that it was individualized. As it was not actually personalised this suggests that there is more than one way to look at personalised/customised learning.

Instead of designing pedagogy and trying to find the missing elements to make the pedagogy work (or supporting it with complementary classroom activities), it is suggested that I have found generic principles on a given subject that not only allow the pedagogy to work but allows the learner to feel that the learning is personalised. McGee (2003, p6) advocates the use of deeper learning principles to develop learning: "If learning objects are designed to support progressively complex knowledge construction, they must be designed around principles that are known to build intellectual capital". Although not the primary intention, on completion of finding the generic principles above it was wondered if these could be used to foster deeper learning as well as aid educators plan pedagogically sound LO learning. Further research is required.

8.5. Model Development

In 2000 little systematic information existed concerning impact evaluations of E-learning in general (Anderson et al 2000). The Capitalisation Report (Leonardo da Vinci programme) identified the lack of *systematic* evaluation as being the major weakness in e-learning projects and in 2003, Hughes and

Attwell (2003) identified the need for robust models. Hughes and Attwell (2003) purport that there has been very little systematic research into the generation of transferable models. Since 2003 there have been developments in this area but very little still exists today concerning potential generic evaluation frameworks (Calbraith 2006). In 2007 I felt that usability/observation of technology interaction was integral to any LO learning model and there were also other considerations. Longmire (2000) believes that research into the pedagogical considerations for the use of LOs requires attention to the design of both LOs themselves and their use within the broader instructional context (Bannan-Ritland et al 2002, Wiley 2003). This highlights the need for effective models that consider the learning environment.

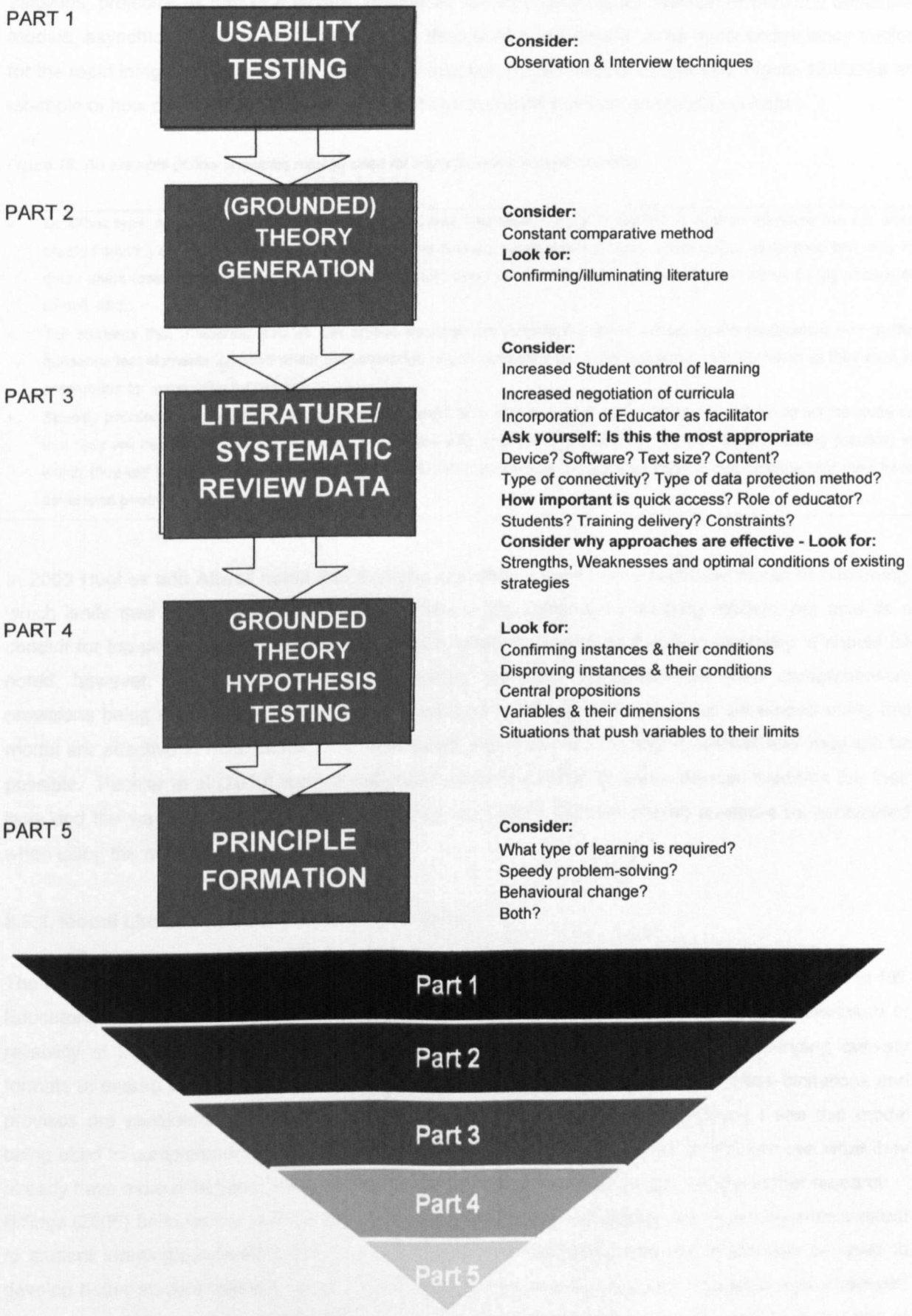
It was obvious that a model could be built using the five component parts of this thesis: Part 1 Usability testing; Part 2 Grounded theory; Part 3 Systematic review of relevant literature (in this example, ICU mobile learning); Part 4 Grounded theory hypothesis testing; Part 5 Generic principle formation. Thus Educators wanting to build a thoroughly rigorous LO Learning environment (customised to their needs) can use this model building either from scratch or by using information already gathered. Replicating my thesis would be ridiculous due to Educator time constraints; however they can use the condensed points located to the right of the model's grey boxes in Figure 16. Waycott et al (2005) found that some users adapt tools to their everyday preferences and Clough et al 2007 stress the importance of incorporating this into mobile learning design. The model therefore takes account of these.

8.5.1. The Model

The model was performed as in Figure 16. Re-ordering the parts are possible but the model's reliability is 'purer' when the order shown is followed (no bias is introduced because GT is done before literature). Several uses of this model are possible: i) Educators can use it in its entirety by replicating the whole process (recommended when creating a completely new LO learning environment on which to base many courses/LOs for distance learning).This is the most rigorous choice but takes the most amount of time; ii) Educators can use this model by simply answering/fulfilling each structure point in turn (located to the right of the grey boxes).This is the least rigorous choice but also the quickest; or iii) Educators can use parts of the model according to what information has already been gathered for course/LO learning environment development. For example, if the full model is desired and the Institution has already conducted usability studies (which address similar questions regarding navigation, platform, aesthetics, and student preferences as these here) the model can be used from Part 2 onwards.
Equally, if the Educator has previously conducted GT hypothesis testing and requires generic principles to form several new LOs, they can either just use the 'main structure points for Part 5' or go through 'Part 5's whole process' depending how quick/rigorous the LOs need to be.

If previous parts of this model have been achieved before undertaking Part 5, Part 5 should provide reliable and context-relevant principles with which to create effective, practice-based LOs for the required learning context.

Figure 16: My LO Learning Environment Model



Once generic principles are gained, this model may have the potential to be used (irrespective of the delivery format) for multiple contexts providing that the objective is the same: e.g. via student computer sessions, projector as part of a lecture, on-line as part of student exam revision or distance education module, asynchronised as part of continued professional development, or as quick competency guides for the rapid integration of new staff members. However, more research is required. Figure 18 shows an example of how principles may be used for an asynchronised distance education package.

Figure 18: An example of how principles may be used for asynchronised distance learning

- **Q: What type of learning is required?** In this example, increased clinical reasoning is desired therefore the LO once created could i) draw attention to important parts of the learning (using flashing lights, bright colour, underlined text, etc); ii) guide users towards reasoned answers (using feedback, direct information, etc); iii) keep attention active (using change of stimuli, etc);
- Tell students that i) working through the on-line package will provide formative self-assessment/evaluation due to the questions/test elements included which aid knowledge retention; ii) they can do this package as many times as they want in preparation for summative evaluation;
- **Speedy problem-solving? Behavioural change? Both?** Both of these are desired therefore the LO could tell the students that they will be expected to partake in an on-line scenario (that combines scientific knowledge and nursing practice) in which they will have to i) prioritise care using knowledge presented in the package; and ii) demonstrate that they have developed principles for practice.

In 2003 Hughes and Attwell noted that systems are often 'locked into a particular model of e-learning' which limits their transferability. My model is free of any particular e-learning models, but acts as a conduit for top-performing LO approaches thereby placing no such limit on transferability. It should be noted, however, that despite generic and unique principles being possible (and comprehensive provisions being made during the research process to ensure generic principles developed using this model are effective in most cases), it is anticipated that 'individual-institution mismatches' may still be possible. Pachler et al (2010) warn about potential disconnection between the way students live their lives and the way educational institutions interact with them. This too should therefore be considered when using the model.

8.5.2. Model Limitations and Necessary Provisos

The model is currently unwieldy if no source data is available or the educator wishes to use it in full. Educators must also beware that strict adherence to the model's suggestions must be undertaken or reliability of the robust approaches will change. Care must also be taken when changing delivery formats to ensure that the same level of approach rigor is maintained. Providing these limitations and provisos are considered/adhered to, the possibilities appear endless. In summary I see this model being used to complement rather than replace existing resources i.e. helping educators use what they already have more effectively. However, full limitations are not yet known and require further research. Billings (2005) believes the real issues revolve around learning technology use regarding *improvement to* student learning/academic programme outcomes. It is suggested that this model can be used to develop better student learning by structuring LO learning in a rigorous and enjoyable way. However, Adams et al (2009) warn "good teaching and engaged learning should not be determined by the use of certain instructional tools but by the guiding principle that learning is an active and recursive process

where knowledge must be contextualised to be relevant to the learner". Therefore the model itself should not be the *overriding* feature. The learner and context should be foremost when using the model. Anything that appears to contradict learner aims/context should be treated as 'suspect'. In the ICU sample, learners said that LOs created in this way made them excited and motivated them to learn more. The LO pedagogy should not only deliver all that it needs to with regard to information but it should be relevant, immediately useful to the learner, and guide their learning to develop informed reasoning.

Theory

According to Glaser and Straus (1967 & 2009) there are two types of theory: i) Substantive (applies to one context with all topics related to the area under study); and ii) Formal (applies to multiple setting or describes the context around the settings. It is compared to all topics, not necessarily related to the area under study). The grounded theory produced in Chapter 1A from original source data represents substantive theory. Exploration of further substantive areas then followed in Phase 2/further format testing, showing that formal theory is possible. Glaser (2006) defines 'Formal Grounded Theory' as a theory developed from several substantive areas, which Glaser and Strauss (2009, p81) believe is necessary to take into account all contingencies met in all the diverse areas that it will be applied. At this stage, attempts to use my theories 'as is' are likely require modification by other theories/comparative analysis (and could therefore not make trustworthy predictions). However, despite generic principles being extracted to form the 'working model' it does not achieve full formal theory status because it does not fully take into account *all* the contingencies, qualifications and/or conditions for its use. In defence of this point, it was not this thesis' primary aim to develop and test a formal theory but to uncover the underlying mechanisms to produce some effective pedagogical and evaluation strategies for learning objects. However, with more extensive field work and testing, this work clearly has the potential to become 'fully formal', and hence a closer examination of this point is warranted.
Glaser and Straus (2009, p237), believe that formal theories should have the following 4 properties: i) the theory should closely fit the substantive area in which it will be used; ii) must be understood by laymen; iii) must be sufficiently general to be applicable in a multitude of diverse daily activities; iv) it must allow the user partial control over changes that inevitably occur during daily use to ensure that it is worth using.

Firstly, my theory fits with the area that it will be used in (i.e. medicine and nursing).

Secondly, although the foundations of this theory have been delivered via conferences (and have been understood and well-received by those who are not familiar with the area), the *full* theory has not been published yet (in press). I am awaiting comments from the wider public concerning parts of the theory incorporated in the model. Therefore, at this time, reports concerning the full theory are not possible and to claim that the *full* theory has been totally understood by laymen would be a little 'previous'.

Thirdly, due to the generic principles, the theory is sufficiently general to be applicable in a multitude of diverse daily activities (evidenced by the effectiveness of theory parts when testing different formats).

Fourthly, the model appears to allow partial user control over inevitable changes wanted or required. Admittedly, it is a little unwieldy in its present form so it is too early to claim that it is a user–friendly asset for educators at this early stage. In summary therefore, it should be noted that although this thesis' theory contains all four properties of a good formal theory, some of the properties need further work for the theory's user comments to become fully known (and therefore refute, consolidate or confirm the theory's real everyday worth). It is noted that theorizing in this manner tends to assert 'generality of scope' (Zetterberg 1963, p52-56) and 'unbounding relativism' (Glaser & Strauss 2009, p68), and that this stance to more open to being disproven than proven by others (Glaser & Strauss, 2009, p63) which is not my intention. The theory produced by this thesis is middle-range i.e. it is not merely a set of hypotheses but is also not a grand, all-inclusive, over-arching theory either. Glaser & Strauss (2009, p30) state: "the researcher's job is not to provide a perfect description of an area, but to develop a theory that accounts for much of the relevant behaviour". Therefore I do not postulate any confident all-embracing theories at this point, but merely suggest that this thesis has highlighted and explained some very important factors, and perhaps may form part of a formal theory.

Conclusions

The exciting discovery of generic grounded theories and methods provides infinite possibilities for development. The challenge for LO learning over the next decade and beyond is enormous. The key to success appears to be in finding methods and models *such as this* with which to systematically assess practice/practical research, and to find approaches that build pedagogically secure foundations for the new e-learning curricula of the future. This will provide a robust and practical basis against which approaches can be evaluated and developed. Hooker (1997, p20) notes: "two of the greatest challenges our institutions face are those of harnessing the power of digital technology and responding to the information revolution". The ability of these methods to adapt to the ever-changing pedagogy (that new ways of learning will undoubtedly require) is paramount.

9. OVERALL INTEGRATION OF CHAPTERS AND FURTHER DISCUSSION OF MAIN THEMES AND THEIR SIGNIFICANCE

Each chapter so far has included its own discussion. There is now a need to pull all chapters and discussions together in order to appreciate their significance as a whole. This chapter will therefore discuss and integrate the main themes further (i.e. LOs, pedagogy, evaluation, evidence base, etc.), revisit my underlying assumptions to estimate whether undue bias has taken place, and revisit the methodology used. It will also address questions raised by each chapter (using the same question phrasing that has previously been used to enable the reader to compare the narrative). My unique and original contributions to each theme/point made will be highlighted throughout in order to illustrate where this thesis 'sits' in relation to the body of knowledge. The chapter ends with a summary of the four most *important* original findings, the limitations of this work, implications for practice, and the need for future research.

9.1. Main Themes

9.1.1. Learning Objects in General

The opening chapter raised several questions (with the intention of revisiting them post-research to provide fuller answers): Will LOs revolutionise learning? Is a better definition of LOs needed? Does 'LO Learning' now have a more reasoned definition? Which parts of Adult Learning Theory are effective for LO learning and result in i) increased user participation or adaption to change? ii) Ability to deal with problems and making reasoned decisions in unfamiliar situations? iii) Adopting a more universal or holistic approach? These will now be addressed.

Will LOs revolutionise learning? Some believed that RLOs would revolutionise learning but this has not yet come to pass. It would be a very naive person to believe that this thesis's LOs could single-handedly revolutionise learning but my work clearly does have the potential to set down original and reliable research-based foundations for LO design, use and evaluation. Feldstein (2006) states that unless a learner actually learns (and at least some of the underlying mechanisms by which this occurs are uncovered) the learning is unlikely to be learner-centred. This is a key point. The learning should act as a 'cognitive catalyst' for the user and thought should be given as to how learning is applied. Both learning (particularly clinical reasoning) *and* its underlying mechanisms were discovered in the LOs designed for this thesis via learner input – the LOs therefore represent effective 'tried and tested' learner-centred learning.

Is a better definition of 'LO' needed? An amalgamated LO definition was used for this thesis. Some (i.e. Hodgins 2002) would argue that there should be a reference to reusability in the definition but I do not agree. It depends on i) the intended/actual use; ii) how reusable an LO may be, should be or needs to be. My unique contribution in this respect is that I have shown LOs as viable and effective learning aids, *able to operate independent of any thoughts concerning reuse* thus no changes to the definition in this respect are required. Bearing in mind that my research highlighted the importance of

design, there may be a greater argument for design to be included in the definition. For example Howard-Rose and Harrigan (CLOE 2003) believe LOs are integral to course design and claim that extra work is created if the design is not integral. 'A little more work required for LO construction' seems a flimsy, unconvincing argument to include design in the definition. In practice, there is mounting pressure on educators to teach greater numbers of students – 'work pressure' would be a more convincing argument. My unique contribution to the LO definition is that my thesis has shown that the *balance of components* is integral to effective pedagogy/evaluation (see Phase 2) which provides the strongest and most credible argument. It is therefore suggested that the word 'any' in the definition should be replaced with the words 'well-balanced', 'carefully constructed', or 'effectively designed'.

Does 'LO Learning' now have a more reasoned definition? Initially this definition was: 'learning that addresses one clearly identifiable topic or learning outcome in any digital resource (reusable or otherwise) that may be used for education/training in any course/curriculum'. This was somewhat 'meagre'. Both this thesis (and indirectly, Feldstein 2006) emphasise the importance of LO learning being 'learner-centred'. Feldstein's reasoning also suggests that LO learning should be measurable. Another obvious and important omission in this definition is the *manner* in which the learning outcome is addressed. My unique contributions to the definition of LO learning are that I have firstly established the term 'LO learning' (it did not previously exist pre-thesis); and secondly I established that a *fully* successful and effective way to design LO learning is either to build on rigorous research-based methods and/or design the learning with fully integrated pedagogy/evaluation approaches. Therefore a more fuller definition of LO Learning is now offered as: 'measurable and learner-centred learning that addresses one clearly identifiable topic or learning outcome in any digital resource (reusable or otherwise) that may be used for education/training in any course/curriculum supported by rigorous research-based methods and/or learning design with fully integrated pedagogy and evaluation strategies, methods and/or approaches'. Furthermore, according to LO participant users, the LOs designed for and used in this thesis had *all* of Robertson & Fluck's (2004) attributes. They were well-received and evaluated by all but one of the preliminary work participants, and by all participants in the main research testing. As this was such a resounding success, it seems fair to say that Robertson and Fluck's description was apt, and that the learning observed in this thesis constitutes 'good' LO learning. My unique contribution here is therefore that I have provided original evidence to support using Robertson & Fluck's (2004) attributes to aid 'good' LO learning.

What parts of Adult Learning Theory are effective for LO learning and result in increased user participation or adaption to change? The scenario in this thesis was able to accommodate all different types of answers offered during the personal development of clinical principles and thus had the ability to adapt to and participate in change. The specific parts of adult learning theory were those contained within top-performing approaches. The LOs were also adaptable to different learning environments without any noticeable affects.

What parts of Adult Learning Theory are particularly effective in dealing with problems and making reasoned decisions in unfamiliar situations? Due to the guided feedback within the LOs learners were able to make reasoned decisions and modify these if incorrect. The scenario then enabled learners to transfer knowledge into unfamiliar situations. The structure rather than certain

parts of adult learning theory in this case were the vital aspects. Specifically, the key structures were: 'providing guided feedback throughout', 'transferring knowledge', and 'personally developing principles'. Research has shown potential for increases in leadership skills and professional confidence when students are provided with "reinforcement of core knowledge and evidence-based information in real-time as required by the student" (White et al 2005, p152).

What parts of Adult Learning Theory are particularly effective in identifying strengths/ weaknesses and undertaking appropriate remediation (i.e. self-directed learning) when compared to other 'non-LO learning' teaching and learning strategies? Self-directed learning was evident throughout the LOs and the scenario to a large degree. Its effectiveness depended upon the LO navigation structure. Where navigation was intuitive this fostered strong learner beliefs that the learning was in line with the personal learning preferences. When searching the literature a quote from Sims (2006, p6) was found to support this finding: "When considering learner preferences it is not a case of designing for an assumed learner predisposition, but ensuring that the learner is able to use their personal preferences to contextualise their experience". Because learners were able to contextualise their learning one of the LOs' greatest strengths was that learners were then able to self-evaluate their progress and identify 'learning gaps'. This gave them 'learning confidence' and clear vision of the required future learning direction resulting in confident goal-setting. Therefore the question is not 'What parts of Adult Learning Theory are useful for learners to identify their own strengths and weaknesses?', instead we should be asking 'Does the LO navigation structure develop self-directed learning?' As top-performing pedagogies (e.g. Chickering & Gamson) produce a higher degree of good navigation (which fosters a very high degree of independent learning) they perform extremely well in comparison to some other general teaching and learning strategies i.e. it is the *structure and sequence* of the pedagogy that are important.

9.1.2. Pedagogy

One broad aim of this thesis was to find out what pedagogical strategies are effective in practice and why. My thesis has been successful in discovering these and has also identified reasons why barriers have been encountered in the past. Previous to this thesis the 'linear', 'deconstruction' and 'context-bound' methods used by Deepwell (2002), Boyle et al (2002) and the New London Group (2000)/Mills (2006) respectively all met with pedagogical barriers and did not adequately address why further development was unforthcoming. The research conducted for this thesis produces several theories to account for these problems.

My evidence suggests that Deepwell's linear method may have chosen an inappropriate starting point, or combined previously unrecognised 'incongruent' pedagogical/evaluation approaches. It is suggested that when using unmatched or untested pedagogical/evaluation approaches concurrently it is probable that either the pedagogy or evaluation (or both) will stop 'working' at some point, thereby providing a barrier to further development. This explains why the linear method of LO construction could only go 'so far' regarding pedagogical development: picking approaches with 'intuitively good face vaildity' is therefore akin to picking them 'out of the air'. It stands to reason that sometimes this type of method may work (in that it accidently may hit on a rigorous method) and other times it may not.

Considering the students 'value-added perceptions' found by the grounded theory, the sum of the parts of Boyle et al's deconstruction/reconstruction method again are equal *greater* than the whole. Postman's (1992) conjecture gives a further clue: New methods change everything, they do not merely add something. Where LO learning is concerned the mere 'addition' of pedagogy to LO tecnological content may cause pedagogical barriers to arise. Instead, content needs to be an *inherent part* of the pedagogical structure. When this is the case, this research has shown that pedagogical barriers do not arise and pedagogies combine easily with evaluation strategies. Thus, it is *the mixture of factors working and interacting together within effective approaches* which, in turn, precipitate implicit value-added learner perceptions to be made explicit.

This also provides another explanation as to why Deepwell's attempts to develop pedagogy did not reach their full potential. In Deepwell's case the value-added learner perceptions were unknown and therefore there was no way of knowing how well it would perform, or how problems with the pedagogy could be resolved or 'dismantled' once encountered. Furthermore, it explains why the New London Group's (2000) and Mill's (2006) 'multi-literacies' pedagogy was effective until it was used as a linear hierarchy or in distinct stages. Reconstructing the pedagogy in these ways may have adversely affected the mixture/balance of components within the pedagogy, thereby affecting inter and intra-pedagogical interactions. Other reasons are also possible: i) there may be effective components within the multi-literacies pedagogy; ii) multi-literacy is effective as a pedagogy but cannot be used in linear hierarchies or distinct stages for reasons as yet unknown; iii) evaluation needs to be considered simultaneously regardless of the pedagogy used. With regard to whether simultaneous development of pedagogy with evaluation approaches is crucial or not, in *practice* this was clear. Pedagogies and evaluation approaches were tested independently of each other (Phase 2) yet both were seen to be effective in different environments, locations and delivery formats. This suggests that pedagogy and evaluation need not necessarily be developed concurrently providing that pedagogical, evaluation and component balance is not destabilised. More important is the fact that the approaches chosen i) contain intrinsic worth (see following section); and ii) complement each other (i.e. they do not cause pedagogical or evaluation problems when using together). As previously stated, when content is an inherent part of the pedagogical structure evaluation approaches are easily combined. I therefore assert that when LO pedagogy/evaluation is designed according to this thesis' recommendations 'methodological synergy' is created.

To this day, researchers still seem to have difficulty locating an adequate starting point to research LO pedagogy, thus my original contributions are as follows: i) It is the *mixture of content (*working and interacting together) *within* effective approaches, which then renders implicit value-added learner perceptions explicit; ii) It is this process which informs theory (as evidenced by the visual paper-based plotting of grounded theory); and iii) due to the latter two points I have supplied both starting points and reference points for further pedagogical research.

Why this research? Putting aside the great need for this work, this thesis discovered approaches that provoke deep learning within an LO context (which in Felstein's words 'create artefacts that function like enzymes for the intellectual digestive system' 2006). This is evidenced by the increasingly motivated response of the learners tested, the deep thought produced, the participants' explicit expression of tacit thinking, and participants being enabled to develop clinical reasoning. LOs can become a seriously effective method for today's learning environments (and accusations against LO

use can be refuted) when i) all caveats have been fully researched; ii) when careful thought is given as to how the LOs are constructed; iii) when careful thought is given as to what the student is actually learning and how LO Learning is applied/delivered; iv) when suitable pedagogies are used; and v) when explicit thought is given concerning how deep thought will be provoked in the learner. Thus the rationale for this research has been supported by this thesis' results showing that LOs constitute potentially good learning for today's on-line learning environments when used with top-performing pedagogical/evaluation approaches. The reasons are two-fold: i) Top-performing pedagogies in this thesis showed a capacity to provoke deep thought; and ii) It is the *pedagogy* used that determines whether deep thought is provoked *not* the LO format per se, e.g. learners were able to form previously unconsidered underlying principles for their clinical reasoning. When using poorly-performing approaches they were unable to do this. It is therefore likely that these findings are entirely due to the approaches used. This links very closely to the following question.

Do LOs have intrinsic worth? It is possibe that intrinsic factors (other than element balance) may also be at work which alter LO effectiveness (i.e. evaluation, IDT, delivery format, etc). As all effective pedagogical approaches appear to work with all effective evaluation approaches regardless of the LO used this suggests that the 'effective factors' are not intrinsic to certain LOs, but are intrinsic to the pedagogical/evaluation approaches. The intrinsic educational worth therefore appears to lie *within the effective approaches* and *not* the LOs. When LOs are constructed correctly, they are however capable of being effective pedagogical/evaluative 'vessels'. Significantly, my thesis has developed the notion of intrinsic worth. My original contribution is therefore that I have been able to confirm the existence of intrinsic educational worth, decipher whether LOs have it, and discover *where* the intrinsic worth lies. My original belief was 'LOs may have intrinsic worth' which then developed into 'perhaps it is the pedagogies/evaluation approaches used with LOs that have the intrinsic worth'. This finally became 'top-performing pedagogies/evaluation approaches have intrinsic worth when used with LOs'. I had assumed that the relation between intrinsic worth and LOs was *passive* (i.e. intrinsic worth is processed by LOs, See Table 46). According to Hjørland's (2005) semantic relation definitions, this assumption proved in reality to be an *active* relation between *pedagogy/evaluation approach* and *intrinsic worth* which in turn *affected* the *LOs*. This is important - the slight difference in emphasis makes all the difference to how the LOs behave which subsequently has impact upon potential uses.

Table 46: Horland's (2005) Semantic Relations

Relation	Definition
Active relation	A semantic relation between two concepts, one of which expresses the performance of an operation or process affecting the other
Associative relation	A relation which is defined psychologically by the people/ Causal relation A is the cause of B
Hyponym us relation	A hierarchical subordinate relation
Locative relation	A concept indicates a location of a thing designated by another concept
Passive relation	A relation between two concepts one of which is affected by an operation or processed by the other
Paradigmatic relation	A relation between two concepts that is established by nature, self-evident or by convention
Temporal relation	A relation in which a concept indicated a time or period of an event designated by another concept

LOs therefore can be effective 'delivery vessels' for pedagogically sound learning. This supports Boyle's general stance that LOs can be pedagogically 'rich', however it should be noted that they can

also be pedagogically 'void' depending onthe approach used. Hence, my original contribution to this area is the discovery and delineation of the relationship between LOs and intrinsic educational worth. This has subsequent 'knock-on-effects' for reusability. Morales et al's developed GLOs due to the fact that the smallest element within the LO was the LO itself which was found to be too big to reuse effectively. Granularity hampered progress so they consequently broke down LOs into their constiuent parts (i.e. content) and focused on content reuse within a generic template. In this thesis, effective pedagogies/evaluation approaches form the generic principles *regardless of content*. This has benefits *over and above* the Morales method which has a *locative* relation (i.e. the LO indicates *where* the content should be because it is designated by the generic template). Their design essentially limits the content to certain locations within the LO. By contrast, the method outlined in my thesis is an *active* relation – i.e. effective pedagogies/evaluation approaches affect how the content is presented. Not only does this render the LO potentially useful *whatever* the format, but there are no such locative limitations on the content. Another unique contribution is that I have supplied effective approaches for educaotrs to use. With regard to choosing an effective pedagogy/evaluation approach for LO learning, if educators use top-performing approaches found by my thesis, they are no longer *restricted* to LOs founded on linear, intuitive or pure deconstruction methods nor on those founded on locative relations.

Do LOs contain intrinsic *potential* concerning their *flexible capability?* As both delivery formats evaluated well it could be said that top-performing pedagogies/evaluation approaches had intrinsic flexibility. However, this again is obviously due to the approaches rather than the LOs because pedagogical/evaluation patterns were seen (i.e. some approaches consistently scored at a similar level despite context). If the intrinsic flexibility witnessed was due to the LOs alone, different levels of flexibility would have been seen during main testing depending on which LO was used with which delivery format.

How far has this thesis achieved pedagogical integration with learner-centred environments? This thesis has shown that 'effective pedagogy' is synonymous with 'learner-centred environments'. Top-performing pedagogies created the learner-centred environment due to its constituent parts, thus integration was inherent.

To what extent has this thesis realised complex pedagogies within complex systems? Sims (2006, p6) implied that *integration* of pedagogy with learner-centered environments is essential. When this thought is married with Rohse & Anderson's ideas concerning the complexities of both pedagogy and the systems involved it becomes clear why pedagogical integration in the past has not been easy: "The literature is replete with recognition... that our notions of learning must extend beyond the psychological processes of individual(s) to one that recognises... complex systems.... Yet there is a sense that this potential to realise complex pedagogies is mostly unmet" (2006, p82). This thesis has therefore succeeded in these respects as all top-performing pedagogical approaches were effective and found to have intrinsic pedagogical worth despite being tested in different forms, disciplines, and delivery formats. Explaining specifically *how* I achieved complex pedagogies within complex systems is a little more difficult to explain. As a result I turned to complexity theory to aid reflection. Complexity theory "is an attempt to understand systems that cannot be explained using the reductive methods of traditional science" (Rohse & Anderson 2006, p83). However, this did not expose anything except the

suggestion that there may not be any direct or easily traceable 'cause and effect'. The 'patterns approach' was then considered as it appears to i) reflect the characteristics of complex learning; ii) recognise the need for flexibility and therefore adaptation, and iii) offer possible solutions when ideas are non-linear and unpredictable; iv) provide "a common ground for researchers, practitioners, technologists and learners... to understand, interpret, evaluate, and share educational practice" (P89, Rohse & Anderson, 2006). However, this did not provide useful insight. Furthermore, Rohse & Anderson (2006, p90) state: "Used within a design-based research framework, design patterns offer a means to incrementally improve education practice". As the patterns approach appears valuable this begs the question: Would this approach have been better for the thesis? As patterns are by nature 'prescriptive blueprints' the difficult part would be finding a pattern that allowed adequate 'approach-housing' flexibility since it is the approach that contains the educational worth. It is likely that several patterns would have been needed for each different type of use for each different subject/location/delivery/discipline. There is therefore no guarantee that the patterns approach would have realised complex pedagogies. Finally simple 'user involvement' was considered to help express specifically how I achieved complex pedagogies within complex systems. Sims (2006, p4) emphasizes the importance of considering users and states that many design strategies, proactive modelling, and interactive metrics fail to address online pedagogy because users are often excluded from the design process. Yen et al (2010) supports this believing that user feedback is critical for evaluating LOs significance. User input was crucial in this research - its pedagogical success was partially due to the fact that users found the LOs so relevant. User observation and users' value-added perceptions gave specific insight specifically into how people learn using LOs and how they engage with LO learning. User involvement aids learner achievement of personal, customised learning in LO design by making the learning relevant (thereby successfully realising complex pedagogies within complex systems). User-informed design should be considered to inform LO pedagogy.

In summary, my unique contribution is the discovery that user involvement is much more helpful (than many patterns, design strategies, proactive modelling, and interactive metrics) when explaining exactly how complex LO pedagogies within complex systems can be achieved.

Does LO pedagogy require a new approach? This thesis suggests that LO pedagogy does *not* require a new approach in as far as some *existing* pedagogies were found to be rigorous and 'top-performing'. As effectiveness appears to rest upon the components *balance* this suggests that approaches should *not* be used in part. Adding/subtracting components may 'destabilise' the LO unless this has been researched or are known to be effective in part. I stated previously that IT had traditionally taken a Constructivist approach to research versus education's more narrative approach. This thesis shows that the IDI model and an all-inclusive type of SR work well together. However, this alone is insufficient evidence to claim that 'established IDT approaches mixed with new systematic (education) approaches *always work* for LOs'. Further testing of all IDT/constructivist/educational approach combinations must first be performed.

Under what circumstances should approaches be flexible/sensitive to technology? It is suggested that approaches should be always be sensitive to technology under all circumstances. Pedagogies should take into account what is technologically possible. Top-performing approaches had the flexibility for content to be presented in many different ways thus avoiding problems.

How do effective approaches inform theory? Phase 2 testing highlighted that this was initially a difficult question to answer however approaches informed theory by identifying what the 'value added' learner perceptions were. Multimedia cognitive theory may also explain why learners believe 'learning is not in line with personal preferences' when timing of information and layout is 'wrong'. Table 47 shows several cognitive assumptions (Mayer 2001, p44):

Table 47: Cognitive Assumptions as Defined by Mayer (2001)

Cognitive Assumption	Description	Authors
Dual channels	Humans possess separate channels for processing visual and auditory information	Paivio (1986); Baddeley (1992)
Limited capacity	Humans are limited in the amount of information they can process in each channel at any one time	Baddeley (1992); Chandler & Sweller (1991)
Active processing	Humans engage in active learning by attending to relevant incoming information organising selected information into coherent mental representations with other knowledge	Mayer (1999c); Wittrock (1989)

The dual channel assumption suggests that when text and images are presented together there is a higher probability that overload will result as these use the same learning channel. Once this channel is 'full' the limited capacity assumption would apply. If too much information is given too fast, learners cannot actively process the learning. Layout observations can also be explained by Mayer's (2001) 'spatial contiguity effect'. When corresponding text and images are adjacent both retention and transfer are improved (the learner is able to hold them in the working memory simultaneously). When text and images are presented too quickly, Chandler and Sweller's (1992) 'split-attention effect' results (forcing learners to mentally integrate disparate sources of information). Thus pedagogical design does not *have to* incorporate all learning styles, *providing that* timing of information and layout is 'good'. This 'good mix of elements' increases the 'active' learning level by allowing the learner to digest the learning. Hence, my unique contribution is the discovery that greater emphasis should be placed on well-timed layout and maximising the value-added principles during the LO design phase.

How do effective approaches inform practice?' In Phase 2, LOs were found to inform practice by i) giving greater/lesser emphasis to different parts of the approaches allowing researchers to identify theoretical gaps and educators to estimate how appropriate the approach is); ii) encouraging learners to prioritise care as in real-life; iii) helping learners to combine both knowledge and practice and increase their critical reflection (greater learning 'linkage' results in greater patient care in practice). Effective approaches that allow learners to self-evaluate enables them to judge learning progression, giving them control over learning, increased 'learning expectations', and confidence. When approaches contain features that allow learners to elaborate on their learning, 'transferable' learning increases.

Are new methodologies needed? My research showed that new methodologies were *not* needed, but new *applications* of the methods *were*. The approach proformas/running orders ensured that all LO learning requirements were met and tailor-made cut-off points for scoring achieved. My unique contribution is therefore that I have succeeded (by addressing the major pedagogical development barriers) where others have failed. I have also provided evidence that the IDI model can be effectively combined with a more 'progressive' type of systematic approach for LO learning; and more importantly that a new approach is not necessarily needed *providing that* the right mixture and balance of pedagogical/evaluation components are assembled.

Does present LO theory constitute a robust basis for LO learning? If looking outside of this thesis, the answer would be 'no'. This thesis identified three authors that unsuccessfully attempted to address pedagogical constraints, thus robust LO theory could not be developed. Overall, this thesis has provided some very valuable pedagogical discoveries, answered some very difficult pedagogical problems and created 'in-roads' into LO learning and how this behaves under certain conditions. My original contribution is that I have been able to bridge the insurmountable divide between pedagogical barriers and LO theory. The result is robust LO theory on which others can build.

9.1.3. Evaluation

To what extent does LO learning need 'non-traditional' evaluation approaches? Surely the most important thing is not whether the evaluation approach is traditional, but whether it is theoretically appropriate and statistically rigorous in practice? It also depends upon the definition of 'traditional'. If this term is placed within education, the MCQs used in this research would be included as traditional methods of summative assessment. As the MCQs were very successful, it could be said that traditional methods can successfully be integrated into LOs. If the term traditional is placed within the context of LO learning alone, it could be argued that this field is too 'new' to have traditional methods. The answer therefore depends on the definition of traditional.

Scriven's (1980) flexible evaluation approach may have subconsciously influenced my thinking regarding the summative assessment created. My method was both rigorous and successful from both educator and student points of view, but Scriven has formative materials becoming summative as a new evaluation phase begins. LO assessment could conceivably be used in the same way with i) time allocated for students to revisit LOs during an allotted formative time period (where answers are recorded and student improvement monitored); ii) The formative LOs could be used both for the formative period and summative assessment (delivered by secure 'locked down' computers if using for distance learning, or with computers and invigilators under exam conditions for site-based assessment. Thus 'knowledge rehearsal' would be possible. However, I did not adhere to Scriven's approach - I use formative work as a 'springboard' to develop transferable skills in the summative scenario. The formative material therefore does not *become* the summative work, but is a *foundation* for it.

Scriven's approach was not fully adhered to thus its potential bias does not actually negatively influence any of the effective evaluation outcomes. However, the inclusion of Scriven's approach definitely shaped my thinking when forming the scenario, meaning that this work is *positively biased* towards Scriven to some degree. Consequently, it is recognised that there may be other ways that may be equally good for summative evaluation but are as yet untested within the context of LO Learning. To gain a fuller answer to the question 'To what extent does LO learning need evaluation using non-traditional approaches?' I am presently undertaking work regarding distance learning formats (separate to this thesis) as they have to by nature 'stand alone'. Thus any important flaws in the LO learning will be very evident in this delivery format.

How flexible do approaches need to be? Under what circumstances? This thesis has gone some way to show how flexible the educational process needs to be. Evaluation approaches need to be flexible enough to cope with today's expansive *choice of technologies* (given the number of access methods and devices students currently use to access information), but also flexible enough to cope

with the *demands of the educator* regarding multiple delivery formats. This ranges from blended through to more 'traditional' learning. They need to be flexible to both these demands in varying degrees depending on what is being taught and how. Specific *optimal* circumstances for these flexibilities require further research. Sheard & Markham (2005, p367) stated: "evaluation of any web-based learning must encompass not only the educational process but also the process associated with the functional usability of the technology". This thesis illustrates how flexible the educational process needs to be. The functional usability of the technology was indirectly addressed through usability studies which showed that when usability aspects are corrected the educational aspects were effective. The model featured in the further format testing chapter should be tested to see how it performs under different conditions at further sites with different populations.

9.2. Other Themes

9.2.1. Evidence Base

It is tempting to ask: **'What is the best way to form an evidence base for LOs?'** however at this stage this is a little premature due to the lack of 'mature' knowledge and practice in the field of LO learning. A more serviceable question would be **'What should LO evidence be based on?'** This thesis has shown that the systematic review (SR) method provides a good way to map out effective LO practice boundaries and grounded theory development added further credibility to this. However, it should be noted that although successful, this thesis presents only *one* such way of mapping out the boundaries. It can therefore not be categorically stated that this is the *best* way just yet as it is presently the *only* way in existence. This thesis has given rise to the possibility that evidence based on tried and tested rigorous pedagogies and evaluation approaches can be developed. As already indicated, more research needs to be conducted testing these types of LOs in different environments, under different conditions in order to *fully* 'map out' the length and breadth of this method's potential. As these LOs were tested on doctors and nurses they had the added theoretical complication of setting appropriate benchmarks in order to comply with current 'fitness-to-practice' documents. However, in practice it was easy to include appropriate benchmarks, to see where grading should sit, and structure the grading accordingly (with regard to evaluation using top-performing approaches).

Do the systematic reviews (SRs) positively establish SR methodology within LO learning (i.e. to aid the testing of quality)? With the inclusion of the caveats outlined in the SR section, the answer is 'yes'. Of course, when considering this methodology the trade-offs between rigor and academic maturity (and the effects/ramifications that these may have) must always be considered. My unique contribution is that I have provided robust SR methodology making SR research for LOs possible.

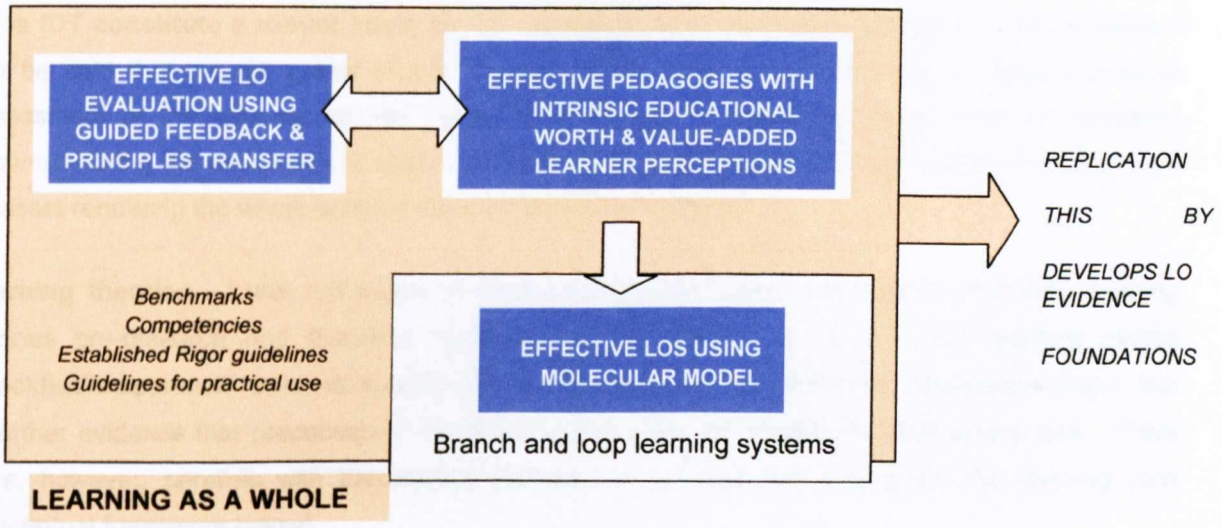
Did the SRs have potential for education/educational research within e-learning? The SRs were clearly useful in drawing out the most important papers. Regarding delivery formats, further format testing was done for 'mobile learning' and there is no reason why many different topics could not be used in the same manner. My unique and original contribution is that this thesis discovered potential for the academic/practical rigor regarding effective LO learning to be pushed higher. It has also highlighted that the inherent rigor of specific approaches have the potential to be increased when the

LO 'field' has developed further, suggesting that future publications on LO research pedagogy/evaluation can, and should, be submitted under this thesis' review protocol so that the developing maturity of the field can be assessed.

Did the SRs have the potential to establish a much needed robust evidence base? Given the success of the main research testing, the methods used provide clear promise concerning a very robust evidence base on which to build. Nevertheless, all the usual caveats for SR use must be considered. My unique contribution is that this thesis provides an alternative to educators having to rely on 'gut feelings' and/or encountering pedagogical problems half way through development.

Where does the Evidence base for LO learning within healthcare 'sit?' At the start of this research there was nothing to base any assumptions on. It was only possible for me to describe the evidence base in terms of *'sitting between three areas'*. Now, instead of having to draw parallels between other topics to explain its status and existence, my unique contribution to this area is that this thesis has developed the LO evidence base to the stage where it can now be described *in its own right* (see Figure 18).

Figure 18: LO Learning's Evidence Base



9.2.2. What Instructional Design Theory (IDT) Should be Used?

What little evidence existed pre-thesis was concerned largely with LO theory and IDT in relation to *reuse*. It was clear to me that some form of instructional design should be used in my research due to the learning's digital nature, and it was theorised (pre-testing) that the IDI model was the Instructional Design Theory of choice (Figure 2). In the initial stages of this thesis the choice of IDT was openly described as 'intuitive'. It was therefore anticipated that upon discussion of the results this choice would be open to discussion.

The decision to use this model as a general guide does not seem to have had any negatively influence/bias on the research but a 'bias assessment' should be made. Since the IDI model was not uppermost in my mind throughout this process, and since development of the thesis was not deliberately or consciously mapped against the model during development, the amount of positive bias *should be minimal*. As the model is built on generic principles it does not appear to favour one

approach over another. When comparing the structure of the model to my thesis, no compatibility modifications were identified or necessary. This showed great alignment between the thesis and the model without deliberate adherence. When examining this alignment further, I noted several important aspects. Firstly, the model allows in-depth assessment of the learning context (in the 'Define' stage) before consideration of objectives ('Develop' stage). This gives educators confidence that learning will be relevant to learners. Secondly, the model renders optimum outcomes self-evident, thereby giving learners 'goal clarity'. Thirdly, the model is easy to use reinforcing my view that educationalists could use it in their own LO learning contexts. Fourthly, the model is capable of being 'wide' enough to allow LO learning development whilst simultaneously being 'narrow' enough to hold the process together. Finally, it does not force component elements into a certain 'mould', nor does it allow the research to become compromised or unfocused by its addition. It makes the LO learning process explicit and fits well with top-performing pedagogies/evaluation approaches. This model is therefore an appropriate and compatible method for LO research and practice and is recommended for this type of work (i.e. LO environment/mobile delivery development).

9.3. Reflections Upon 'Underlying Assumptions' and Potential Bias

Does IDT constitute a robust basis for LO learning? IDTs were not specifically tested, however it can be said that the IDI model proved to be a robust basis for LO learning. It allowed in-depth assessment of the learning context *before* consideration of objectives. As a result no problems occurred during the development stage. The emphasis on process and greater depth of testing were an asset rendering the whole process clear-cut and uncomplicated.

Learning theories - I was not aware of having any preconceived assumptions regarding learning theories pre-research and therefore no bias was anticipated. As my personal learning beliefs (Brookfield's approach) were not reflected at all in the types of top-performing approaches found, this is further evidence that preconceived learning theories have not biased the work in any way. There were, however, parallels with Brookfield's approach in the way that I designed the learning (see Conceptual framework below).

Theoretical framework - I was not aware of having any preconceived assumptions pre-research except perhaps that a usable theory on LO learning may be possible. Not only has this thesis shown that a usable theory *is* possible but it has also shown that it is *effective when used practically*.

Conceptual framework - Conceptually I acknowledged the presence of a personal approach similar to Brookfield's. The level of *good learning conditions* actually displayed in the learners during testing will be examined in order to estimate the level of influence Brookfield had. During the research learners described the learning as 'very relevant', 'customised', and 'authentic' implying that the students perceived the learning as being related to their own experience. It was later found to promote ownership and deep learning due to its structure. The LO topics, (as described in the surveys preliminary chapter), were deliberately designed to help students deal more effectively with their everyday problems (i.e. deterioration of patient condition due to an increased understanding of underlying pathophysiological and medical/nursing management principles) as evidenced in

summative scenario scores. The LOs demanded direct practical application of both the knowledge and principles learned from the users. Although increased excitement, interest, and motivation for learning was evidenced by LO users, the active research data-collection 'window' was not long enough to be able to adequately assess or evaluate whether the LOs helped learning to be seen as 'enhancing job satisfaction and self-esteem' – nor was it adequate to make any judgements concerning increased 'Lifelong Learning mentality' of users. Brookfield also states that the learning conditions are *most* effective when the learning incorporates elements of challenge to promote critical analysis taking learners beyond what they already know (e.g. promotes the transition from 'information' to 'knowledge', or 'a little knowledge' to 'greater knowledge').

The research in this thesis provided a 'critical analysis challenge' in the form of active reflection parts within the LOs, followed by parts where learners had to apply the critical reflection. There is potential for critical reflection to be an even greater part of this process. For these reasons, I am presently researching (beyond this thesis) incorporation of a critically reflective diary to be used as part of the 'formative LO use' period – where learners are required to catalogue and appraise critical incidents that have occurred in practice in relation to what they have learned in theory from the LOs. Summative marking of these reflective diaries will be undertaken (alongside the summative LO scenario) using an amalgamation of a host university marking scheme and a research-based critical reflection marking scheme (see Calbraith 2001). Brookfield's final effective learning condition is when the learning incorporates 'the needs of the organisation and society as well as development of the individual'. My thesis research was sensitive to these needs *particularly* during further format testing. These findings will therefore be discussed in depth with regard to the organisation where they were used, society and the individual learners:

Organisation:

i) The primary stakeholders were consulted pre-research (i.e. Director of Nursing, Nursing Standards Officer, Ward Manager, Lead Clinician/Clinical Director, Head of Teaching and the Teaching Team). This enabled me to identify where the LOs would be most needed with regard to the testing unit (i.e. helping newly qualified nurses 'get up to speed' quickly). The LOs not only helped newly qualified nurses towards increased clinical skills progression, but more senior nurses found them useful physiological 'revision' methods.;
ii) Consulting stakeholders helped to ensure that the LO terminology would be familiar to the learners therefore bearing less risk of 'format/language alienation';
iii) Research data collection time was deliberately delayed for four months in order to be sensitive to the increased workload of the nurses induced by a geographical move of the unit.

In summary, the LO learning considered the needs of the organisation and was rewarded by a smooth running data collection period, but most importantly, valid and effective learning took place.

Society:

The needs of society were not researched *per se,* however society's need to have competent doctors and nurses is self-evident. If they are not well-trained, then it stands to reason that patients may be put at risk of mistakes, misdiagnosis, mismanagement, and late detection of what could have been preventable causes of death and morbidity.

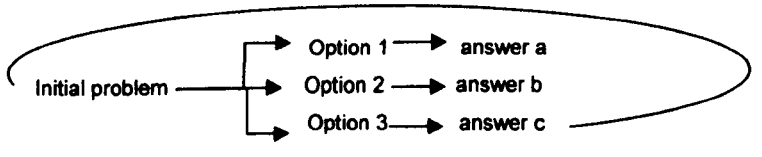
Individual Learners:

The LOs were effective and sensitive concerning the development of the individual learner. Each learner without exception showed increased skill at using clinical reasoning (as evidenced by fewer initial wrong answers as they worked through the LOs. Each learner was also expected to transfer learned principles under summative conditions, therefore each learner had no choice but to apply these due to the nature of the LO assessment structure.

As virtually *all* of the 'conditions that influence learning' were displayed in the course of the research this suggests that Brookfield's approach has shaped my thinking to a large degree. More importantly, 'To what extent does this constitutes negative bias?' Significantly, Brookfield's approach did not have any direct links or similarities to the top-performing pedagogies/evaluation approaches - it is therefore unlikely that Brookfield's approach has influenced the research unduly. Furthermore, no adverse effects were found within the thesis - the research was sensitive to both individuals, organisations, and indirectly aware of societal needs. As such, Brookfield's conditions are considered to be a large asset to my thinking, rather than a negative bias, and may provide a very useful conceptual framework for this type of work.

Learning Systems Approach - In the introduction Downey's (2003) 'branch' and 'loop' learning systems were mentioned. Pre-research, I had an underlying assumption (based on experience) that LO learning materials require both 'branch' *and* 'loop' learning systems and therefore assumed that a *new mixture* of each type (where the branches ultimately connect back onto themselves forming a loop) are both possible and desirable (See Figure 19).

Figure 19: Assumed Branch and loop learning



To create a null hypothesis, the LOs had been designed so that the learner could choose their own route through the learning material in order to shed more light on this assumption. Participant route was observed and recorded in detail during fieldwork observation/usability studies. The results are important and very interesting, and are best described here using examples taken from the usability studies. I had rightly assumed that both branch and loop systems were needed. For each part of the LO 'back' and 'home' buttons were desired/used. I had anticipated that the branches ultimately would connect back onto themselves forming a loop but assumed that they would be 'isolated components' that repeated wherever revision of the question/answers or learning was needed (i.e. as in Figure 19). The main reason for believing this was that I knew participants often want to look back at learning to consolidate it further. However, I did not anticipate the two main routes observed when participants were looking at *more than one* aspect within their chosen topic (see Figures 20 and 21):

Figure 20: Actual Branch and Loop Linear Route Taken

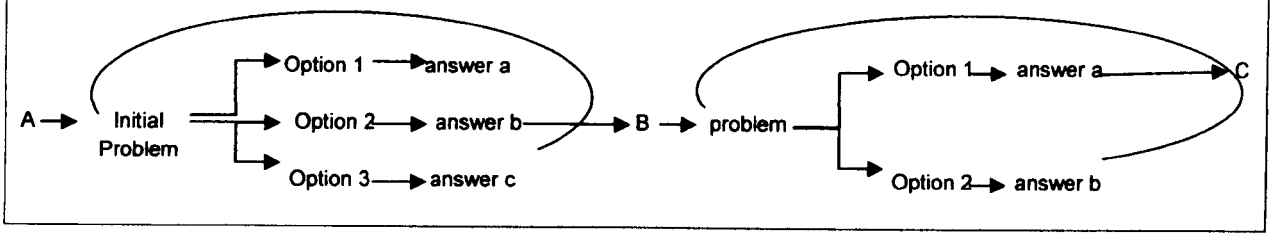
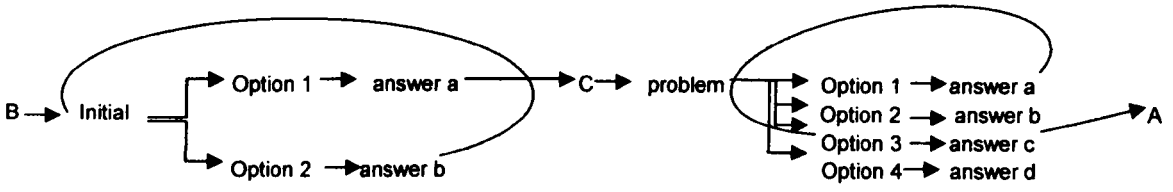


Figure 20 shows the linear route taken by *some* of the participants. Figure 21, however, shows an example of the random type of route taken by *most* participants:

Figure 21: Actual Branch and Loop Random Route Taken



This is a very important finding. The main difference between 'what had been observed' and 'what was anticipated' was that I had expected participants to use loop learning on each of the problems (A, B, C) *separately* - i.e. the loop would be complete before going onto the next question/problem/learning part. I had also expected them to complete each answer before wanting to revisit information. Although this did happen on occasion, I observed that loops were happening before questions were completed. When asking participants the reason for their choices it became obvious that loop learning happened for two main reasons: i) when participants were checking information/concepts they were not certain of; and ii) because participants found it easier to learn this way (learning preferences). This explains why the routes in Figure 21 were the predominant ones - i.e. without being asked to, *participants were customising their own learning* by revisiting information when *they* chose to. This further adds weight to the argument that both branch and loop learning are necessary for effective LO learning.

Specifically, there were some differences between the disciplines but none of these affected the fact that branch and loop learning was being performed as and when the learner chose for as many times as the learner desired. Women in the 'Laypersons group' in the preliminary usability work were the only group to use a linear route through the LOs without exception. This suggests that when information is very new or 'foreign' to the learner, the learner (and perhaps women in particular) will tend to use a linear 'branch' type of learning on the initial run through, and then loop learning when revisiting the material. Thus one reason for the students' subconscious choices may be linked to confidence – i.e. the newer the information the lower the confidence which in turn affects the route taken through the learning. When considering Mezirow's 'transformative learning' and Benner's 'novice to expert' theories it is obvious that there are 'cut-off points' where learning and knowledge integration becomes automatic, thereby transforming the learner into a professional and competent person. In this case, the

perceived lack of confidence is likely to persist until 'transformation to autonomous practitioner' has taken place (Calbraith 2001).

Previous experience led me to believe that learners need to choose what they learn in order to maximise the relevance (hence the need for different 'learning branches'). However, during the learning process learners often requested further 'back' buttons to be added to the package when asked about improvements in order to revise chosen sections. This suggests that packages with purely linear branches are inadequate and further underlines the need for 'learning loops'. This is perhaps evidenced by differing student scores (linear route people scored less compared to loop route people). Both the learning systems required and the routes taken through the learning are very important as they indicate i) how effective learning can take place, and ii) the level of confidence of the learner (regarding how many times they need to check information). When collected, this information can be used to modify questions that students routinely find 'too hard' for example. Thus when trying to encompass i) learners of all levels of 'learning confidence'; ii) learners with all learning styles and preferences; iii) customised learning and iii) adequate critical reflection (to enable sound clinical practice to develop into research-based competencies and accurate clinical reasoning), both branch and loop learning should be present in the LO learning. In summary, the assumption that LO learning requires both branch and loop learning was not only correct, but it also provided the springboard for me to develop original theory on LO Learning systems (as described above).

Philosophical approach - The overriding approach taken for this thesis was largely pragmatic. The philosophical standpoint had to be one that would encompass mixed methods methodology in order to obtain both what worked and why. Traditionally, the mixture of instructional VLE/e-learning material and didactic teaching methods have met with low alignment expectations (please see Figure 8). As a result, efforts were taken during this research to improve alignment of both learners and educators in the form of i) asking learners and stakeholders (usability studies) what they would find the most useful in their working/learning environments (i.e. social constructivism); and ii) giving guided feedback at regular intervals to learners in order to develop clinical reasoning (i.e. facilitative on-line teaching). High alignment expectations were further reinforced by some of the blinded pedagogy research (e.g. Chickering & Gamson's 1989 'Good learning principles'). It did this by mentioning verbally that learners would be able to achieve a high mark due to its structure. Learners randomly assigned to this evaluation method scored marginally higher than other methods as it seemed to increased learners' confidence and expectations that the learning would be relevant and easily attainable.

At a very basic level, my awareness of Powell's ecology of e-learning philosophies influenced the research only in the sense that high alignment expectations were maintained. The low degree of overall influence may be because I did not have a high affinity with Powell's approach. As there are no obvious direct or indirect influences it is perceived not to have negatively biased the work. As the ecology was not very enlightening post-research it is suggested that a more comprehensive philosophical continuum may be better for this type of work.

9.4. Revisiting Methodology

The mixed method approach methodology was proposed as the best method for this thesis. This decision is now revisited by examining exactly how component parts behaved and how effective they were.

9.4.1. 'Systematic Reviews' (SRs) Versus 'Case Series/Ethnography'

The decision to reject a simple literature review is upheld as SRs showed all potential outcomes, to what level of rigor they had been evaluated to, and their impact on practice. Since the SRs rendered i) explicit method descriptors (thus others can replicate then to test, verify or refute the claims of this research); and *ii)* explicit pedagogical/evaluation approaches (allowing generic principles to be formed), this suggests that I was correct to reject the narrative review format as these aspects would not have been so obvious or possible. Even so, some may question my choice concerning SRs due to the narrative nature of some of the data, and the grounded theory 'revelations'. However without SRs, specific evaluation/pedagogies would not have been identified so easily (there would have been no 'what' to hang the 'why' on without considerable amounts of work beyond the scope of this study). There would also have been no clear overall baseline provided, no geographical mapping of LO learning, no estimation of where this field is at regarding academic maturity, nor an evaluation of the quality of reporting. Some authors believe that only SRs containing Randomised Controlled trials (RCTs) are true SRs. The fact that the research types included in this thesis' SRs proved to be both 'workable' within the review format *and* in practice, suggests that this format was appropriate. This approach also established whether scientific findings were consistent; whether they could be generalised across populations and settings; whether findings vary significantly by particular subsets (where possible); and the reasons for heterogeneity. It generated new hypotheses about particular groups/original themes; suggestions for selecting and justifying starting points for research; and minimised some of the bias encountered in both small and very large trials. It stated how meta-analysis increased the precision of the estimates and established a way to provide information to not just the specialist but also the non-specialist using 'short' versions of the potential model. As such I refute the belief (where LO learning in concerned) that only RCTs should be included in SRs.

Concerning avoidance of SR pitfalls, there were three points of note: i) Many educational confounding variables (e.g. unforeseen curriculum changes) could make some SR variables very difficult/impossible to measure. Due to my data collection sheet design, study objectives, materials and methods were made explicit ('transparent' in their clarity of design, and in the way they are reported and analysed). This ensured that many of these types of confounding variables and SR pitfalls were avoided by rigor sheets discarding those studies; ii) When new papers published from this date onwards are submitted to the SR protocols (and subsequent changes made) care should be taken to estimate whether the LO model remains responsive and adaptive to evolving evaluation needs. Although slightly cumbersome, it must be remembered that this model is in its infancy. In development terms the model represents 'early inception'. It should also be remembered that there is presently no other method/model in existence that is responsive in this way; iii) there is basic need for research to continue. Although, top-performing evaluation approaches encompassed not only the educational process but also the process

associated with the technological functional usability, and although they were capable of being fused, dismantled, and/or re-fused with other pedagogies, SRs proved to be indispensable. They provided a rigorous evidence base for LO learning and demarcated exactly 'what' was working in practice with regard to research-based pedagogy/evaluation. This said, it does not replace the need for basic research to continue.

As expected the SR method did not fully elucidate concise reasons as to 'why' the approaches were effective, but did however provide comprehensive factors to consider and be compared with grounded theory in each case. The grounded theory provided useful hypotheses which proved to be largely generic when tested in Phase 2/Further format testing. This not only provided reasons as to why SR factors were important but also illuminated relationships *between them*. Potential bias from 'unscrutinizable' source data was prevented by using new data from usability studies. Technical faults were minimised/corrected quickly and therefore did not become extraneous variables. Consequently deductive and inductive formats were highly compatible. Furthermore, if the compatibility of deductive and inductive formats were to be judged purely on terms concerning whether questions that my thesis posed were answered, they would be compatible.

Some may argue that interviewing should have been used instead of usability studies as this is more 'in-keeping' with overall inductive methods. Several potential advantages of this are noted: i) time required for the research may have been reduced if interviews alone had been conducted; and ii) the 'ethnographical sympathies' of the surveys fit well with the grounded theory method (Glaser & Strauss 1990, p35). However these possible advantages would not have occurred in reality. It is now clear that i) time saved not having usability studies would have been required later to fix technical problems; ii) a difficult evaluation would have been required to see whether findings were due to the research or extraneous variables; iii) an evaluation would also have been required to see whether possible extraneous variables had had an adverse effect on the research; iv) additionally, as usability studies proved to be unexpected sources of *rich* data, the time saved not having them may have been negated by the extra time required for longer interviews to gain equal amounts of information; and v) the inclusion of ethnographical sympathies does not mean that *only* pure ethnography will suffice. Furthermore, Kuzel (1992) advocates 6-8 data units (individuals/studies/sites) for homogenous groups and 12-20 for heterogeneous groups. Thus when comparing sites, case studies would have required at least 6 instead of 2, and to achieve 'typicality' more than 6 would definitely have been needed. This would have given insufficient time to complete them. Having conducted the research, I remain of the opinion that rejecting ethnography/case studies was wise, and taking the usability testing/GT route was both more efficient and 'time-effective'. This is further consolidated by the following reasons: i) Despite taking perhaps the shortest route through theory development, the PhD timeframe was only just adequate. This further underlines that case series were not a viable option. The anticipated danger of this thesis becoming unwieldy (due to large numbers of separate cases needed to show any commonalities if case series had been chosen) appeared an accurate prediction. Hence using SRs and usability studies was not only a theoretically valid decision, but also a pragmatic one; ii) the usability data and grounded theory proved more than adequate for the purpose; and iii) major ethical mismatches experienced between the chosen two sites created problems that were extremely difficult to remedy. This meant that adding further sites would further complicated these difficulties and would

probably have compromised overall research generalisability. These compromises may have become so great as to render typicality unfulfilled and the project unworkable/unreliable - negating its value.

As always, there is a 'trade-off'. Undoubtedly, slightly 'richer, lived experience' *may have* been captured using ethnography, as Rohse and Anderson (2006, p88) state: "Those who have applied ethnographic methods to find out what learners actually do when presented with resources and opportunities for self-managed e-learning have found learners' lived experiences to be richer and more complex than following the prescriptive pathways anticipated by their teachers". However, this would have been at the cost of less generalisable and possibly incomparable data which may not have resulted in generic principle production. The data captured using usability/GT proved to be rich enough to produce effective, reliable and valid principles. It was also rich enough to not only explore and theorise concerning LO learning but also how the technology aids or interferes with that process. It was therefore deemed to be a successful and robust method to use. This is supported by Ryan and Bernard (2003) who state that textual data with verbatim text without rich narratives is sufficient to enable theory to be discovered. Interviews enhanced this process further by consolidating the verbatim text with 'richer meaning'.
Using '5 users plus' in one round of testing proved to be more than adequate to find all technical problems/desired modifications. This was supported by no further new problems discovered during actual research testing (There was only one occasion during Phase 2 when a thunder storm produced a power surge temporarily 'freezing' the screen. Seconds later the user was able to continue).
In summary therefore, under *normal* circumstances, usability testing was able to identify/rectify all possible faults/technical 'distractions', thus extraneous variables were not left undetected and therefore did not impact upon testing approaches later on.

When correcting problems, discarding Krug's method in preference for Garrett's proved to be wise decision and a useful tool when considering to LO modifications. Usability studies enhanced the student experience by i) helping learners to see that the technology was aiding the learning process; ii) making learning 'context-appropriate'; and iii) including learners in the design process. The usability findings uncovered even richer data than expected and were therefore a greater asset than even anticipated. Furthermore, findings were easy to assimilate into the grounded theory process to produce measurable hypotheses. Additionally, I gained a greater understanding of i) how people learn; ii) how learners can optimally be engaged in online environments; and iii) the importance of 'integrated' pedagogical environments. Overall, the usability studies were much more than just an appropriate method to gather the source material required for hypothesis testing.

To conclude, the particular mix of methods used for the thesis stood the test of time and rigor, and produced excellent researcher understanding. On these grounds, the decision to opt for a mixed method approach is upheld.

9.4.2. Grounded Theory

It is not presently known whether there are optimum types of LO source data regarding grounded theory (GT). This thesis did not specifically research optimum types but as previously stated the

combination of usability, observation and interview performed well. Playing devil's advocate, there is an argument that just taking source data from only two sites for main testing was limited and unlikely to represent the whole population. If the blinded results of the main research hypothesis and approach testing had not found the very similar results as the initial SR/grounded theory observations I would have to agree. The fact that all results (gained through different methods and different populations, locations and disciplines) show similar findings indicates that more generic or intrinsic principles are at work. As the SRs considered all studies presently in existence this provides an argument that all rigorous studies within the population have been incorporated. It is duly noted however, that having performed the grounded theory in 2 samples there is a danger of subconsciously looking for similar codes or interpreting data under those codes. It is important to consider my own reactions and motivations with regard to the data to estimate to what extent this is true and whether it constitutes a major flaw in this work. It is true that initially I thought generic principles may be possible; however I felt genuinely astonished to find that samples and particularly codes were *so* similar in different samples. Despite putting forward optimum pedagogies/evaluation approaches in Phase 2, I was genuinely surprised when blinded participants *repeatedly* rated the top-performing approaches as 'great learning' without me speaking at all. It is therefore not likely that my interpretations have introduced major flaws into the research (as development as always been based on 'pure' verbatim participant comments) emphasising the validity and reliability of the findings. Even though my findings are strongly indicative of being generalisable to the general population of medicine and nursing, I choose not to confidently assert this until others have replicated this work.

Regarding the alternative GT method (i.e. gathering all known source data in existence), it is likely that the anticipated time/workload required in reality would probably *not* have taken any more time/work than what I actually spent on three SRs and two usability studies. This may suggest that using grounded theory alone for this type of study is an equally valid and pragmatic approach as the one used here when considering time issues only. However, other issues are important too. Clearly, extricating case co-ordinator bias would have been difficult unless co-ordinators were observed.

Secondly, SR data was occasionally 'sketchy' making full categorisation impossible. The level of ambiguous data may have increased if 'already existing' data was used, and may have compromised the findings.

Thirdly, to produce an evidence base worthy of medical profession 'acceptance' would have required 'gold standard' case studies to be produced necessitating further time commitment. Also no 'absolute' estimate of bias would be possible (from the learners' answers) without identifying them, interviewing them, and making estimations concerning the honesty of their answers. No 'absolute' estimate of researcher bias would be possible either without interviewing them too (This is providing access/permission were given and that learners could be married up with their answers given during others peoples' research). Not only does this introduce extreme difficulties but it contravenes research governance and ethical codes where participants' answers are confidential/blinded. Pursuing such a course would also require full data sets in each case if reliable results were to be obtained and bias limited. The evidence base produced would have been fraught with problems when trying to make generalisations. Given all these things, the overall decision to reject 'grounded theory used alone' as the best method for this work is upheld.

It is clear that the particular combination of GT methods used for the thesis stood the test of time, produced 'purer' data and complemented deductive methods well. On these grounds, the decision to opt for a mixed method approach is overwhelmingly upheld. Having examined mixed method component parts, the mixed method approach was an extremely good decision. It can be seen that not only was the general approach effective, but I believe that the specific components parts chosen drew out information that could not have been gathered any other way.

9.5. The Five Most Important Findings

Because LO learning is a new area for Medical Education literature, effective governing principles and working practice have yet to be fully established. As such I have either created original methods (or used original applications of established methods) for virtually every component of this PhD with the exception of the grounded theory method. Thus my PhD is highly original. The five *most important* findings are as follows:

i) This thesis establishes the *intrinsic worth* and *flexibility* of effective approaches i.e. w*hat* evaluation and pedagogical approaches are effective in LO design and *why.* This thesis has shown that educators should choose evaluation approaches that allow the learner to gather information systematically to faciliate the development of individual learning plans *during* the learning process. Similarly, pedagogical approaches chosen for LOs must respect diverse ways of learning and communicate high expectations of achievement to learners. Student input in the evaluation process, immediate guided feedback, intuitive layout, and active learning are all essential elements;

ii) It is the 'value-added' perceptions of learners that hold the key to unravelling pedagogical development barriers - these perceptions are the often unseen and uncaptured thoughts of students as they learn. This type of user input was used throughout and was crucial for understanding unexplained pedagocial barriers in LO development;

iii) Effective generic principles for pedagogy and evaluation *are possible* for desktop and mobile delivery. This thesis has identified generic principles which can be used by educators to create effective LO learning *or* form benchmarks on which to build an LO evidence base e.g. evaluation pinciples include demonstrating how new knowledge is easily applied and how it builds upon existing knowledge to create individual learning plans for learners. Pedagogical principles include reminding the learner what they have learned, identifying how to apply it, and keeping attention active;

iv) LO Learning systems use and therefore need branch *and* loop learning (i.e. learners need to be able to have the choice to progress both in a linear fashion through the material and to look back at it) My thesis has delivered evidence/created a different way of thinking about learning systems in that many observed participants did not wait until they had finished a section or even a question before wanting to look back at their learning. Each particpant did this wherever it suited

them during the learning and as frequently as they wanted. Participants therefore had a much higher-than-expected desire to customise their own learning in this way;

v) LO structure based on Wiley's 'molecular' format evokes excitement and motivation in the learner – i.e. adding or taking away effective pedagogical elements from the LO has the effect of stabilising/destablising the LO. For those with a knowledge of chemistry, this can be likened to adding or taking away electrons from the outer shell of an atom. In simple terms, adding or taking away parts of the learning format can weaken the whole learning and evaluation* structure. When the balnacne of elements are 'right' learners become motivated. In the same way that electrons interact with other electrons/electron shells. It is the interactions between pedagogical and evaluation* components that affect LO learning. Thus it is the relationship between the learning theory and its evaluation that influences how effective the learning will ultimately be. Wiley (1999a) describes his molecular LO analogy from an 'atomic' level, however I believe this theory can be taken even deeper to the 'sub-atomic' level, i.e. electrons can be split into constituent parts (quarks, etc) and so can pedagogical elements. It is at *this level* that the generic principles are found. It is ytherefore suggested that unless we proble deeper into pedagogical and evaluation* approaches, universal principles for effective LO learning will not be fully unearthed.

This thesis has therefore achieved *all* of the broad aims set out in the introduction, *and* has also discovered a potential new learning system and an innovative way of increasing student motivation. The overall limitations, implications for practice and future predictions concerning this thesis will now be made clear.

9.6. Limitations

Most limitations have been discussed throughout however a summary will be given here. The major limitation of this thesis is that the model produced cannot *yet* be used *unthinkingly* due to the need for further reseearch. Careful thought must be given to ensure that whatever version of the model is used is appropriate for the context in question, especially when delivery formats (other than those tested here) are used. It must also be understood that to choose a shorter version of the model means a deliberate choice to use a less vigorous method. This too should be married with the level of rigor required. Widely publicised disadvantages of the SR method were overcome by careful planning and implementation. In relation specifically to the LO model development, some may consider the SR element to be unwieldy - and rightly so. Most educators simply do not have the time to undertake a full SR for their chosen topics thus 'short-cut versions', condensed points and generic principles were offered to make the results of this thesis 'usable' for hard-working time-limited professionals that want or need to build rigorous/effective LO learning quickly (N.B. Because *all* delivery formats were not tested the efficacy of using other delivery formats on these approaches are unknown. It may also be possible that generic principles may not be possible/clear when using other delivery formats hence further research is required).

A second possible limitation involves SR rigor. Some may argue that the 'cut-off' point for inclusion may be slightly low and therefore potentially represents an unacceptable compromise. However, I fervently defend my decision for the following reasons: i) No studies were accepted that should have been rejected. This is evidenced by the quality of the approaches found for Phase 2 testing; ii) Later on in this process when attempting to align the rigor 'cut-off' level with KSF competencies, LO MCQs, and professional 'fitness-to-practice' benchmarks the rigor level aligned easily without any mismatches or compromises to safe practice; iii) This thesis showed that educational SRs just do not *warrant* the same rigor as clinical trials (thus my decision to set the bar at 'the majority of rigor categories should be ticked to be included in the SR' was actually more than sufficient to discern between good and bad quality studies); iv) These LOs are the first of their kind. Initial SRs on any subject are (by the very 'nature of the beast') perhaps a little lower in 'achievable rigor' than those performed several years later due to lack of maturity in the field that increases as time progresses. Indeed, many SR collaborations (e.g. BEME, Cochrane, Campbell) ask reviewers to regularly update the literature/papers included after initial SR publication to ensure that overall conclusions do not merely remain robust as time passes, but also to push the 'acceptable' achievable rigor a little higher. It must also be remembered that top-performing approaches were found by *blinded participants* and a *null hypothesis*.

The most rigorous pedagogies tested in this thesis have shown a capacity to provoke deep thought as users were able to form previously unconsidered underlying principles for their clinical reasoning. Whilst it is true that the *LO format* seems to have also encouraged deep thought, it may not be due to LO learning itself. It is possible that these findings may have been almost *entirely* due to the effective pedagogies and evaluation strategies used. Thus a further limitation of this work is that a more expansive answer on this subject is not possible without doing further testing using different delivery formats with the same effective pedagogies and evaluation approaches (e.g. on-line lectures, podcasts, small group work in collaborative online synchronised 'spaces', etc.). N.B. If these formats also provoke deep thought it may indicate that it is *exclusively* the pedagogies (rather than LOs in any way) that have caused this to occur. What is clear, however, is that LOs (providing they are constructed correctly) are capable of being effective pedagogical or evaluative 'containers'.

Two points of note remain - regarding grounded theory and generalisability. The grounded theory for each population was almost *uncomfortably* similar. If the grounded theory hadn't been blinded and verified over and over again using different popluations and locations it would have raised suspicions regarding preconceived interpretation of core categories. However, as core categories were all slightly different (i.e. not one of the pursuant samples had the exact mixture of core codes each time), and great pains were taken to maximise and minimise conditions. this confirms that interpretations are likely to be unbiased. The best way to evaluate generated substantive/formal theory (according to Glaser and Strauss 2009 p224) is if the theory is "accurate in fit and relevance to the area it purports to explain". As previously stated, the fit was excellent.

This thesis's findings can tentatively be generalised to the wider population as there was a mixture of countries and disciplines with regard to participant and study characteristics (see Sections 5.5.2. and 5.5.3). However, this is *only* in the widest sense of the word as specific countries/disciplines were not deliberately targeted or representative samples sought. When attempting to generalise this thesis' findings to the general population of Nursing and Medicine care should be taken to observe the mentioned caveats (i.e. generalisability is only robust in the *predictive sense* thus this thesis does not claim 'carte blanche' generalisability). Only when others have replicated this work, and when further work on specific representative samples have gained similar robust results can unmitigated reliability of generic principles in all discsiplines, delivery formats, locations and populations be affirmed.

9.7. Implications For Practice

Most of the general implications of my work on the body of knowledge for educators are that they no longer have to rely on intuitive LO pedagogical design/development, but now have comprehensively researched top-performing approaches to use. Specifically this thesis offers this a working model to develop LO learning environments which are capable of being customised to educators' own specifications. This will undoubtedly demand a culture change and a greater facilitative role for the educator with regard to the software/hardware, and helping students become accustomed to it. The implications for students are that they will have LOs that encourage deeper thought, active reflection and development of clinical reasoning. They should also have greater engagement with the learning, and because learners are actively part of the development process (survey/usability testing, etc) they should have greater ownership. In the workplace the changes in organisation and culture necessary to implement this kind of learning may demand unfamiliar decisions to unfamiliar questions: 'Will mobile learning interfere with existing equipment?' It will also demand some of the usual ones: 'What are the benchmarks? Is the learning cost-effective?' Stakeholders too are part of the process outlined in this thesis thus a further implication is that a greater level of 'meeting of minds' (between stakeholders, universities', and the students' desires) may be witnessed. If so, a greater level of satisfaction for all is possible concerning relevant and good quality learning. Implications for web developers on one hand are no different i.e. my approach to LO development did not use any particularly technologically innovative approaches and the LOs described here can be delivered using basic and widely-available programmes. Conversely, the challenge for IT will be in ensuring that any summative assessment is securely delivered if used for distance or mobile learning, and if monitoring of student use/progress is required ensuring this is accomplished.

Post-thesis the main implications for LO learning are that it has a robust basis. Educators can now confidently develop LO learning knowing that certain 'reference points' and 'working principles' have been mapped out (thus hastening further development). Implications for learning as a whole is that students' speed of learning should be faster and greater linkage between theory and practice seen.

8. Future Research Required

It is clear that all initial aims of this thesis were covered and new knowledge added. It also added important indications as to whether LOs can transfer successfully to *all* delivery formats. Due to the largely generic nature of the top-performing approaches, principles and model it is predicted that the possible applications for these are almost infinite. However, because *all* delivery formats were not tested it is also possible that the approaches found may not be *as effective* when using other formats (e.g. as part of inter-professional learning).

Furthermore, generic principles may not be *possible* when using other delivery formats and effective mobile pedagogies may work equally well with desktop delivery. However, these aspects have not yet been fully researched. Effective mobile or desktop pedagogies/evaluation approaches may be appropriate for both delivery formats (i.e. approaches are interchangeable) but if this is not the case mobile and desktop pedagogies should be deemed as *incompatible* and specific *optimal* circumstances for delivery flexibilities would therefore require further research. Having discovered generic principles further research is required to discover whether they represent 'deeper learning principles' as defined by McGee (2003). If so, they may support progressively complex knowledge construction.

As the overall approach to LOs described in this thesis is newly formed (and testing the IDI model was not a primary aim) it is impossible to estimate the level of comparable rigor between this and other models. As such, no *categorical* claim can presently be made as to whether this is the *best* model to use for LO learning. Therefore testing the rigor of many different IDT models using this LO approach is now required. Further research is also required to establish whether greater learning ownership levels are gained using 'self-owned' versus 'loaned' devices when LOs are used during mobile learning. If so, this may have an impact on the depth of learning.

Finally, with regard to the model produced, full limitations are not yet known. Experts have not yet used my model to build new on-line learning environments nor used my generic principles to complement (rather than replace) their existing resources. The principles may help educators use what they already have more effectively (see Table 48). This table shows handy hints and tips for educators when wanting to incorporate the generic principles found in this thesis with an existing on-line learning package or idea. Having chosen the aims, objective and content of the learning educators may consider the following:

Table 48: Aide memoire for Generic Principle Use

Choices and Considerations	Result	Comments
Choose an evaluation* approach that allows the learner to gather information systematically	This allows the learner to self-evaluate and therefore helps them to develop personal learning plans *during* the learning process	
Choose a learning system that allows the learner to revisit previous questions during formative learning periods	This allows the learner to customise their own learning	
Choose a pedagogy that respects diverse ways of learning (i.e. accepts more than one way to phrase an answer/accept it as correct)	This allows the learner to develop their own reasoning behind the correct answer rather than just learning the answer	This does *not* mean that only long answer questions are appropriate e.g. MCQs can be used but should be accompanied with immediate guided feedback when the learners chooses an answer
Choose a pedagogy which also communicates high achievement expectations to learners	This allows the learner to realise that when they have completed the learning they will have achieved the required competency level	
Consider the inclusion of immediate guided feedback	This allows the learner to rethink the rationale behind their chosen answers	
Ensure that the layout of the learning material is 'intuitive'	This allows the learner to navigate easily (and therefore understand where their learning gaps are). Unclear layout leads to feelings of lack of control and learning choice resulting in feelings that the learning is not in line with participant learning preferences, that knowledge desire has not been satisfied, and that learners have not been given the time to 'digest' the learning (overload).	Improving layout clarity makes learner self-evaluation easier and gives learners increased feelings of greater control and learning choice. This results in learners believing that the learning is in line with personal learning preferences *no matter what those preferences are*, and actual time required to digest the learning decreases.
Consider the inclusion of active reflection	This keeps learning active and focuses reflection on specific information that will help the learner to find the correct answer	
Consider inclusion of an element that requires learners to transfer principles learned	This allows the learner to check their reasoning behind why they think their answer is correct	

Undoubtedly this type of use would tease out further limitations and consolidate the conditions under which it can be used. It is hoped that the processes and approaches outlined in this thesis will be practically useful for educators, those developing e-learning environments and those commissioning learning contracts alike.

* The term 'evaluation' refers here to evaluation approaches (as defined and labelled by the published authors). Once an approach has been chosen and embedded into the LO structure it can of course be used for assessment purposes also.

10. CONCLUSION

Great precision of care was taken throughout the whole thesis to ensure that each stage of its development and each component part used was as rigorously constructed as possible. Meticulous rationales for use were sought which I believe has resulted in a cohesive whole and is evidenced in how well the parts have integrated. The succesful identification of effective aproaches was no doubt enhanced by this level of precise care.

Having discovered the most effective pedagogical and evaluation approaches for the design and use of Learning Objects within Medical Education, this thesis has laid out some of the underlying pedagogical and evaluation principles *within* these approaches. It has given educators several references points for development and removed the pedagogical barriers often encountered in this type of work.

This thesis provides an original and excellent basis on which *LO learning* can be further explored suggesting that the inital methods of enquiry were well-chosen and the inductive development stages well-conducted. Working with clearly defined questions, being open to research design, following grounded theory, and undertaking systematic reviews has both *elicited and verified* theory. The methods used here proved to be not only appropriate for this research but they have potential applications for further testing. The systematic reviews and 'value-added learner perceptions' established the boundaries of *intrinsic educational worth* and learner input was crucial in the understanding of underlying mechanisms of effective approaches. A fresh look at learning systems revealed the high level of self-customised learning desired by today's on-line learners. Similarly, scrutiny of Wiley's molecular analogy provided a useful comparison when contemplating 'element balance' during LO design.

This thesis offers original approaches and generic principles which may form possible benchmarks for educators to use in the design and development of their own LO materials and environments. Should the methods used in this thesis become an accepted way of LO teaching and learning, future applications could be compared with benchmarks for IHEP standards (Appendix 41) and different student groups. Such measurements would provide insight (into distance learning student perceptions during their educational process for example) where few accepted measurement methods exist. In conclusion, I believe that this thesis provides some significant contributions to the important area of LO learning.

11. REFERENCES

Adams, N.B., DeVaney, T.A., & Sawyer, S.G., 2009, Measuring conditions conducive to knowledge development in virtual learning environments: initial development of a model-based survey, pp.1-23, available from www.jtla.org

Ades, A.E. & Higgins, J.P.T., 2005, The Interpretation of Random-Effects Meta-Analysis in Decision Models. Medical Decision Making 25, PP646-654. Also available from: http://www.statsdirect.com/help/meta_analysis/heterogeneity_in_meta_analysis.htm

Agostinho, S, Oliver, R., Harper, B., Hedberg, J., & Wills, S., 2002, A tool to evaluate the potential for an ICT-based learning design to foster "high-quality learning". In Williamson, A., Gunn, C. Young, A. & Clear, T. (Eds.), Winds of change in the sea of learning. Proceedings of the 19th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education, Auckland, New Zealand: UNITEC Institute of Technology, pp29-38

Alexander, S., 1999, An evaluation of innovative projects involving communication and information technology in higher education, Higher Education & Development 18 (2) pp173-183

Ally, M., 2004, Constructing and creating learning objects Part II: Chapter 6, In McGreal, R. (Ed.) Online Education using learning objects, London, Routledge/Farmer, p1

ALT, 2005, Research Strategy, Paragraph 11 paragraph 11 (retrieved 09.09.05), Available from: www.alt.ac.uk/ALT_2005_Research_Strategy_20050420.html

Anderson, T. A., 2003, I Object! Moving Beyond Learning Objects to Learning Components, *Educational Technology*, July-August, pp19-24

Anderson, J. & McCormick, R., 2005, Ten Pedagogic Principles for E-learning, Available from: http://insight.eun.org/ww/en/pub/insight/thematic_dossiers/articles/quality_criteria/equality2.htm

Antman, E.M., Chalmers, T.C., Lau, J., Kupelnick, B., & Mosteller, F., 1992, A comparison of results of meta-analyses and randomised controlled trials and recommendations of clinical experts, JAMA 268, pp240-8

Ausubel, D. P.,1960, The use of advance organizers in the learning and retention of meaningful verbal material, Journal of Educational Psychology 51: 5, pp267-272

Ausubel, D. P., 1968, Educational psychology: A cognitive view, New York: Holt, Rinehart & Winston

Bachman, J.A., & Panzarine, S., 1998, Enabling student nurses to use the information superhighway, Nurse Education 37(4), pp155-61

Bain, J.D., 1999, Introduction, Higher Education Research and Development 18 (2), pp165-172

Bandolier, 2010, Available from: http://www.medicine.ox.ac.uk/bandolier/booth/glossary/hetero.html

Bannan-Ritland, B., Dabbagh, N., & Murphy, K., 2002, Learning object systems as constructivist learning environments: Related assumptions, theories, and applications, In D. A. Wiley (Ed.), The Instructional Use of Learning Objects (pp. 61-98). Bloomington, Indiana: AIT/AECT. Available from http://reusability.org/read/chapters/bannan-ritland.doc.

Barritt, C., Lewis, C. & Weiseler, C., 1999, Cisco Systems Reusable Information - Object Strategy: Definition, Cisco's Creation Overview and Guidelines, From: http://www.cisco.com/warp/public/779/ibs/solutions /learning/whitepapers/el_cisco_rio.pdf

Barron, T., 2000, Pioneering applications of learning objects, Learning Circuits, 1 (3). Also available from www.learningcircuits.org/mar2000/barron.html

Bartlett, J.F., 1995, Experience with wireless world-wide web client, Technical report, note TN-46, Digital Western Research Lab., from: www.research.digital.com/wrl/techreports/abstracts/TN-46.html

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., et al, 1996, Improving the quality of reporting of randomized controlled trials: the CONSORT statement, JAMA 276, pp637-639

B.E.M.E., 2000, Best Evidence in Medical Education (BEME: Report of meeting 3-5 December 1999, London UK), Medical Teacher, 22(3), pp242-245

Billings, D. M., 2005, From teaching to learning in a mobile, wireless world (Guest editorial), Journal of Nursing Education, Vol.44, No.8, p343

Borau, K. & Ullrich, C., 2009, Mobile learning with open learning environments at Shanghai Jiao Tong University, China (January–April 2009), Learning Technology Newsletter, Vol.11, Issue 1 & 2, pp7-8

Boskic, N., 2003, Faculty Assessment of the quality and reusability of learning objects (unpublished Masters Athabasca University, Canada; cited in Krauss, F. & Ally, M., 2005, A study of the design and evaluation of a learning object and implications for content development, Interdisciplinary Journal of Knowledge & Learning objects, Vol.1, pp1-21

Boyle, T., 2002, Towards a theoretical base for educational multimedia design, Journal of Interactive Media in Education, Available from: http://www-jime.open.ac.uk/2002/2/

Boyle, T., and Cook, J., 2003. Learning Objects, Pedagogy and Reuse, in Seale, J. K., (Ed.) Learning Technology in Transition: From Individual Enthusiasm to Institutional Implementation.

Brandon Hall, 2009, Mobile learning comes of age: how and why organizations are moving to learning on mobile devices, Available at: www.brandon-hall.com/publications/mobilelearning/mobilelearning-comesofage.shtml

Bridge, P.D., & Ginsburg, K.A., 2001, An integrated approach for evaluating students' achievements of clinical objectives, Medical Education Online, 6:9, pp.1-6, also available at: www.med-ed-online.org

Britton, B.K., Stimson, M., Stennett, B., Gülgöz, S., 1998, Learning from instructional text: Test of an individual-differences model, Journal of Educational Psychology, Vol.90, No.3), pp476-491

Brookfield S.D., 1987, Developing Critical Thinkers. Challenging Adults to Explore Alternative Ways of Thinking and Acting, Buckingham, Open University Press, pp1-4

Brown, S., 2007, Designing For Effective Learning: A Critique of Generic Learning Outcomes, Journal of Learning Design, Vol. 2, No. 2, pp22-30

Burns I. & Glen S., 2000, A new model for a new context? In Eds. Glen S. & Wilkie K. Problem-based learning in Nursing, Basingstoke, Macmillan Press, pp1-8.

Calbraith, D., 2001, Towards a reflective framework and marking scheme, Masters Thesis-M.Med.Sci. Clinical Education, Greenfield Medical Library, Nottingham University

Calbraith, D. 2006, Systematic approaches to E-learning methodology and evaluation, Published Oral paper given at the International iPED conference, October, Coventry University

Calbraith, D., 2010a, Discovering a reliable quality base for E-learning evaluation, Abstract and oral paper published in the RCN 2009 International Research Conference proceedings, Blackpool

Calbraith, D., 2010b, Analysing Student experiences using Touchstone, Abstract and oral paper published in the RCN 2009 International Research Conference proceedings, Blackpool

Calbraith, D. & Dennick, R., 2009a, Pedagogical strategies for learning objects – Published Abstract, Oral paper also given at the ICERI International conference, Madrid, Spain, 16-18th November 2009, IATED publishing, Also available at www.iated.org/iceri2009/index.php?section=19

Calbraith, D. & Dennick, R., 2009b, Pedagogical strategies for learning objects – Published full paper, t Oral paper also given at the ICERI International conference, Madrid, Spain, 16-18th November 2009, IATED publishing, Also available at www.iated.org/iceri2009/index.php?section=19

Calbraith, D. & Dennick, R., 2010, Producing Generic Principles and Pedagogies for Mobile Learning: A Rigorous Five Part model, Chapter 2 in: Kitchenham, A. (Ed), 2011, Models for Interdisciplinary Mobile Learning: Delivering Information to Students, IGI Global, Pennsylvannia

Carmean, C., 2002, Mapping the Learning Space, from http://www.west.asu.edu/nlii/learningmap.htm

Carroll, J.M. (2000) Making use: Scenario-based design of human-computer interactions, p87 also available at: www.mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=4114

Chalmers T.C., Levin, H., Sacks, H.S., et al, 1987, Meta-analysis of clinical trials as a scientific discipline: control of bias and comparison with large co-operative trials, Stat Med, 6, pp315-325

Chehimi, F., Coulton, P., & Edwards, R., 2006, Delivering 3D advertising to mobile phone, Proceedings of the IEEE, International conference on consumer electronics, Las Vegas, January 2006. Cited in Chehimi, F., Coulton, P., Edwards, R., 2006, *Advances in 3D graphics for Smartphones* available at: http://eprints.lancs.ac.uk/22928/1/getPDF17.pdf

Chickering & Gamson, 1989, '7 priniciples for good practice in undergraduate education' cited In Christiansen, J. & Anderson, T. 2004, Feasibility of course development based on learning objects: research analysis of three case studies, p2, from: http://www.itdl.org/Journal/Mar_04/article02.htm

Christiansen, J. & Anderson, T. 2004, Feasibility of course development based on learning objects: research analysis of three case studies, p2, from: http://www.itdl.org/Journal/Mar_04/article02.htm

Howard-Rose, D., & Harrigan, K., 2003, CLOE Learning Impact Studies LITE: Evaluating Learning Objects in Nine Ontario University Courses, (CLOE: *Co-operative Learning Object Exchange,* From CLOE stories webpage), Available from: http://learnware.uwaterloo.ca/projects/CCCO/cloe_stories.html

CNET, 2010, Available from: http://reviews.cnet.com/best-smartphones

Cobb, S.C. & Baird, S.B.,1999, Oncology nurses' use of the Internet for continuing education: a survey of Oncology Journal of Continuing Nursing Society Congress attendees, Education in Nursing, 30(5), pp199-202

Cochrane review, 2010, Available from: http://www.cochrane-net.org/openlearning/HTML/mod13-3.htm

Cohen, E.B., 2006, Learning Objects and E-Learning: an Informing Science Perspective, Interdisciplinary Journal of Knowledge and Learning Objects, Volume 2, p24

Cook, D.J., & Mulrow, D., 1997, Synthesis of best evidence for clinical decisions, Systematic reviews: Annals of Internal medicine, 126; pp376-380

Cooper, I. & Shufflebotham, R., 1995, PDA Web Browsers: Implementation issues, pp.1-12, Available from http://alethea.ukc.ac.uk/wp?94ihc

Cramphorn, C., 2004, An evaluation of the formal and underlying factors influencing student participation with e-learning web discussion forums, Paper presented at the Network Learning Conference, April 5-7, Lancaster

CRD (Centre for Reviews and Dissemination), 1999, Stage II Phase 7 (chapter 6 & 7) CRD document, CRD, Putting Research into Practice, Effective Healthcare (6)1

CRN, 2010, Informing Science, IS Institute, Santa Rosa, CA, USA, Available from: jsessionid= DTFGA1N1W4DPVQE1GPDA at http://www.crn.com.mobile/222002587

Creanor, L., Trinder, K., Gower, D., & Howells, C., 2006, LEX The learner experience of e-learning: Glasgow Caledonian, University, Glasgow, Final project report

Dabbagh, N., 2000, Re-designing instructional strategies for On-line learning, paper presented at ALN 2000, Adelphi, Maryland, USA, Available from http://mason.gmu.edu/~ndabbagh/wblg/matrix1.htm

Dabbagh, N., 2003, The intersection and alignment of Learner-Centred instructional strategies in on-line and their implementation using course management systems, presentation at NL11 Next generation CMS, Tuscon, AZ, From: http://www.educause.edu/asp/doclib/abstract.asp?ID=NL10330

Darby, J., 2003, Learning environments for student-centred learning, cited in J. Christiansen & T. Anderson, 2004, Feasibility of course development based on learning objects: research analysis of three case studies (second page), Available from http://ets.berkeley.edu/etstandards/ukeu/Darby.ppt

Davis, D., 2005, (May/Jun.) The integration of research and practice: the importance of examining the evidence, From the Research Basics series. Issues related to the integration of research and clinical practice. Neonatal Network, 24(3), pp77-9

Davis, D.A., Thomson, M.A, Oxman, A.D., & Haynes, R.B., 1995, Changing Physician Performance: a Systematic Review of Continuing Medical Education Strategies, Journal of The American Medical Association, Vol.274, pp700-705

Dearnley, C., Haigh, J., & Fairhall, J., 2008, Using mobile technologies for assessment and learning in practice settings: A case study, *Nurse Education in Practice*,Vol.8, No.3, pp197-204

Deepwell, F., 2002, Unpublished Interactive Framework (based on Oliver & Conole's 1998 model using an illuminative approach, countenance model and action evaluation model)

DeSimonian, R. & Laird, N., 1986, Meta-analysis in clinical trials, Controlled Clinical Trials 7, pp177-188

De Ville, K., 2008, The cure is in hand? The brave new world of handheld computers in medicine, *Cambridge Quarterly of Healthcare Ethics* (Cambridge University Press), 17, pp.385-400

Dickelman, G. & Banerji, A., 1999, Concept development theory - PERFORMANCE SUPPORT FOR THE NEXT MILLENIUM: A model for rapidly changing technologies in a global economy; Proceedings of the HCI International '99 (the 8th International Conference on Human-Computer Interaction) on Human-Computer Interaction: Communication, Cooperation, and Application Design, Munich, pp187–191 Available from: (http://www.sigmod.org/sigmod/dblp/db/conf/hci/hci1999-2.html#DickelmanB99). PPT: http://www.pcd-innovations.com/downloads/PS_for_next_mil.ppt

DoH, 1999, Making a Difference: Strengthening the Nursing, Midwifery, and Health Visiting contribution to Health and Healthcare, London HMSO

DoH, 2000, Making a difference: strengthening the nursing, midwifery and health visiting contribution to Health and Healthcare, London HMSO

DoH , 2000b, The NHS Plan: a plan for investment, a plan for reform London HMSO

DoH, 2001, Learning from Bristol: the report of the public Inquiry into Children's Heart Surgery at the Bristol Royal Infirmary, 1984, London TSO

Downes, S., 2000, LOs from; http://www.usdla.org/html/journal/JAN03_Issue/article01.html

Downes, S.M., 2004, The rise of learning objects, http://www.itdl.org/Journal/Mar_04/editor.ht

Downey, S., 2003, Design, Standards and Reusability 31 July 2003, Available from: http://www.downes.ca/cgibin/website/view.cgi?dbs=Article&key=1059622263&format=full

Dron, J., 2007, Control and constriant in e-learning: Choosing when to choose, London: Idea Group publishing

Dumas, J.A., Dietz, E.O. & Connolly, P.M., 2001, (Jan-Feb) Nurse practitioner use of computer technologies in practice, Comput Nurs 19(1), pp34-40

Dumas, J.S. and Redish, J.C. (1999) A practical guide to usability testing Exeter, England, Intellect publishing, p122

Duncan, C., 2003, Chapter 2 from: A Littlejohn (Editor) Reusing online resources: a sustainable approach to e-learning, London, Kogan Page, p13

Eaton, K., 2010, The iPad's biggest Innovation: Its $500 price, Fast Company, Available from: http://fastcompany.com/article/apples-tablet-introduced?page=0%2C0

Education and Skills Committee, 2005, UK e-University: Third Report of Session 2004-05, London: House of Commons, p17

Egger, M., Smith, G.D., & Sterne, J.A.C., 2001, Uses and abuses of meta-analysis, Clin Med JRCPL 1, pp478-484

E.N.B., 1998, Developments in the use of an evidence and/or enquiry based approach in nursing, midwifery and health visiting programmes of education, ENB, From: http://www.enb.org.uk/evidpub.htm

Engel C.E., 1991, Not Just a Method but a Way of Learning, In Eds. Boud D. & Felletti G., The Challenge of PBL Learning, London, Kogan Page, pp22-33

EPPI Centre, June 2002, EPPI-review: A systematic review of the impact of summative assessment and tests on students' effectiveness for learning, EPPI.

Ehrenberger, H., & Murray, P.J., 1998, (Nov-Dec) Issues in the use of communications, Oncol. Nurs., Technologies in nursing research 25 (10 Suppl), pp11-15

Estabrooks, C.A., O'Leary, K.A., Ricker, K.L. & Humphrey, C.K., 2003, The Internet and access to evidence: how are nurses positioned? Journal Of Advanced Nursing, 42(1), pp73–81

Farrell, M. J. & Rose, L., 2008, Use of handheld computers in clinical nursing education, Journal of Nursing Education, Vol.47, No.1, pp13-19

Felder & Silverman (1988). Towards A Comprehensive Learning Object Metadata, In Yahya, Y. & Yusoff, M., 2008, Incorporation of Context to Stipulate Meaningful Learning and Enhance Learning Object Reusability, Interdisciplinary Journal of E-Learning and Learning Objects Volume 4, (Formerly the Interdisciplinary Journal of Knowledge and Learning Objects)

Feldstein, M., 2002, What is usable E-learning? E-learn magazine webpage, Available from: http://www.elearnmag.org/subpage.cfm?section=tutorials&article=6-1

Feldstein, M., 2006, There's no such thing as a Learning object E-learn magazine webpage, Available from: http://www.elearnmag.org/subpage.cfm?section=opinion&artcile=74-1

Felix, U., 2005, E-learning pedagogy in the third millennium: the need for combining social and cognitive constructivist approaches *ReCALL*, 17, pp85-100

Ferneley, E., 2010, Feasibility study into MP3 and MP4 Assisted learning, Available from: www.edu.salford.ac.uk/scd/tlqis/abstracts/ferneley0506.rtf

Fleiss, J.L., Gross, A.J. 1991, Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: A critique. Journal of Clinical Epidemiology 44, pp127-39

Friedman, C., De Bliek, V., Greer, D., Mennin, S., Sheps, N., Swanson, D., & Woodward, C., 1990, Charting the winds of change: Evaluating innovative medical curricula Academic, Medicine 65, pp8-14

Friesen, N., 2006, (January 24) Building a better LOM? ISO Metadata for Learning Resources, from: http://www.cancore.ca/betterlom.html

Ford P. & Walsh M., 1994, New Rituals for Old: Nursing through the Looking Glass, Oxford, Butterworth Heinemann

Gagne, R., 1985, Active Learning Theory - See Gagne, R., 2004, Gagne Home Page, from: http://www.sjsu.edu/depts/it/edit186/gagne.html

Gagne, R., 1992; Active Learning Theory - See Gagne, R., 2004, Gagne Home Page, from: http://www.sjsu.edu/depts/it/edit186/gagne.html

Gagne, R. M., 1986, Instructional technology: The research field, Journal of Instructional Development, Vol.8, No.3, pp7-14

Gagne, R., 2004, Gagne Home Page, from: http://www.sjsu.edu/depts/it/edit186/gagne.html

Ganger, A.C. & Jackson, M., 2003, Wireless handheld computers in the preclinical undergraduate curriculum, Medical Education On-line, 8:3, available from www.med-ed-online.org

Garrett, J.J., 2003, The elements of user experience: User-centred design for the web, AIGA, New Riders, New York, pp19-35

Gavaghan, D.J. et al , 2000, An evaluation of homogeneity tests in meta-analysis in pain using simulations of individual patient data, Pain 85, pp415-424

Gibbons, A.S., Nelson, J., & Richards, R., 2002, The nature and origin of instructional objects in Wiley D.A. (Ed) 2002c, The instructional use of learning objects (chapter 1.2) Bloomington Indiana AIT & AECT 25-58

Gilbert, J., Morton, S., & Rowley, J. 2007, E-learning: The student experience, British Journal of Educational Technology 38, No.4, pp560-73

Glaser, B., 1998, Doing Grounded Theory: Issues and Discussions, Mill Valley: Sociology Press, pp16-9, p187

Glaser, B., 2003, Teachable Moment, Paper presented at the Grounded Theory Research Seminar, London

Glaser, B., 2005, A Teachable Moment. Paper presented at the Grounded Theory Research Seminar, San Francisco.

Glaser, B.G., & Straus, A.L., 1967, The discovery of grounded theory, Aldine, Chicago

Glaser, B.G. & Straus, A.L., 2009, The discovery of grounded theory: strategies for qualitative research, Aldine, New Jersey,US

Glasziou, P., Vanderbroucke, J., & Chalmers, I., 2004, Assessing the quality of research, BMJ 328, pp39-41

GMC, 1993, Tomorrow's Doctors: Recommendations on Undergraduate Medical Education, London, GMC

GMC, 2002, Revised Tomorrow's Doctors, London, GMC

Goth, C., Frohberg, D. & Schwabe, G., 2006, The focus problem in mobile learning, Proceedings IEEE 4th International workshop on wireless, mobile and ubiquitous technologies in Education (WMTE'06), Athens, pp153-160

Green, S., 2005, Systematic reviews and meta-analysis, Singapore Medical Journal, Vol.46, No.6, pp270-274

Greenhalgh, T., 1997, How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses), BMJ 315, No.7, p109 (Education and debate)

Habermas, J., 1971, Knowledge and human interests, Beacon Press, Boston

Hagan, I., & Markham, S., 2000, cited in Sheard, J. & Markham, S., 2005, Assessment and Evaluation in Higher Education, Vol.30 No.4, pp 353-368).

Haines, A., & Donald A., 1998, Getting Research Findings into Practice, London, BMJ books

Harasim, L. M. (1988). Online Group Learning/Teaching Methods (Technical Paper #7). Education Evaluation Centre, The Ontario Institute for Studies in Education.

HEA, 2006, Taken from: Laurillard, D., 2005, Presentation given at the HEA 2005 conference: Harnessing technology to improve the learning experience, Higher Education Academy.

Higgins, J.P.T. & Thompson, S.G., 2002, Quantifying heterogeneity in a meta-analysis, Statistics in Medicine 21, pp1539-1558

Higgins, J., Thompson, S., Deeks, J., & Altman, .D., 2002, (January) Statistical heterogeneity in Journal of systematic reviews of clinical trials: a critical appraisal of guidelines and practice, Health Services Research and Policy 51, Vol.7, No.1, The Royal Society of Medicine Press Ltd, pp51–61

Hodgins, H.W., 2002, The future of learning objects in: Wiley, D.A. (Ed), 2002b, The instructional use of learning objects (chapter 5.3), Indiana Bloomington AIT & AECT

Holloway, I. & Wheeler, S., 1996, Qualitative Research for Nurses, Blackwell Science, Berlin, Germany, pp89-111

Holloway, I. & Wheeler, S., 2010, Qualitative Research in Nursing and Healthcare (Third Edition), Wiley Blackwell, Chichester, UK, p102

Holzinger, A. Nischelwitzer, A., & Meisenberger, M. (2005) Lifelong learning support by M-learning: example scenarios, Association of Computing Machinery, ACM eLearn Magazine 5, p12

Hooker, M., 1997, The Transformation of Higher Education, Cited in The Learning Revolution: The Challenge of Information and the Academy, Oblinger, D.G., & Rush, S.C. (Eds), Bolton, Mass, Anker Publishing, p20

Howard-Rose, D., & Harrigan, K., 2003, Cloe Learning Impact Studies LITE: Evaluating Learning Objects in Nine Ontario University Courses, CLOE. Available from: http://learnware.uwaterloo.ca/projects/CCCO/cloe_stories.html

Hubble, M.A., Duncan, B.L. and Miller, S.D. (Eds), 1999, The heart and soul of change: what works in therapy, Washington, DC, American Psychological Association Press

Hutchinson, J., 1986, Grounded theory: the method, in Munhall, P.L., & Oiler, C. (Eds), Nursing research: A qualitative perspective, Connecticut, Appleton – Century Crofts, Norwalk, pp111-130

IEEE LTSC, 2002, Draft Standard for learning object meta-data, Learning Technology Standards Committee, p6. Also available from http://ltsc.ieee.org/

Inglis, A., Lung, P. & Joosten, V., 1999, Delivering digitally: managing the transition to the knowledge media, Kogan Page, London, p157

International Skills, 2003, Survey of Instructional Development Models: The IDI model (adapted from K.L Gustafson's model) Syracuse University

Jones, R. & Boyle, T., 2007, Learning Object Patterns for Programming(Learning Technology Institute, London Metropolitan University, London, UK, Interdisciplinary Journal of Knowledge and Learning Objects, Vol.3, p26

Jong-Ki, K., 2008, The effects of LMS quality and e-learners characteristics regarding e-learner's scholastic performance: A proposal for e-learning success model Proceedings from ASBBS 15, no.1 pp34-35

Jurczyk, J., Kushner Benson, S.N. & Savery, J.R., 2004, Measuring Student Perceptions in Web-Based Courses: A Standards-Based Approach, College of Education, The University of Akron, Available from; jpj2@uakron.edu

Kaufman, J., 2006, Practical Usability Testing, p3, available from http://www.digital-web.com/articles/practical_usability_testing/

Kelley, K., Clark, B., Brown, V., & Sitzia, J., 2003, Methodology Matters: Good practice in the conduct and reporting of survey research, *International Journal for Quality in Health Care* 15, pp261-266

Keppell, M., Kennedy, G., Elliott, K. & Harris, P., 2002, Transforming Traditional Curricula: Enhancing Interactive Multimedia Medical Education through Problem-Based Learning, Multimedia and Web-based Resources, Electronic Journal of Computer-enhanced Learning

Koper, E.J.R., 2001, Change to renewal, in: Schramade P (Ed), Handboek Effectief Opleiden 26, Elsevier, The Hague, pp45-86

Koohang, A. & Harman, K., 2005, Open source: A metaphor for e-learning, Informing Science: The International Journal of an Emerging Trans-discipline 8, pp75-86, also available from:
http://inform.nu/Articles/Vol8/v8p075-086Kooh.pdf

Koper, E.J.R., 2001, Modelling Units of Study from a pedagogical perspective: the pedagogical meta-model behind EML, IMS Working Group document, Available from: http://eml.ou.nl/introduction/docs/ped-metamodel.pdf

Krauss, F. & Ally, M., 2005, A study of the design and evaluation of a learning object and implications for content development, Interdisciplinary Journal of Knowledge & Learning objects, Vol.1 pp1-21

Kristoffersen, S. & Ljungberg, F., 1999, "Making place" to make IT work: empirical explorations of HCI for mobile CSCW. In Proceedings of GROUP "99, ACM Press, New York, NY, pp276-285

Krug, S., 2000, The companion Website, p147-159, from: www.stevekrug.com

Krug, S., 2006, Don't make me think: A common sense approach to web usability (Second edition), New Riders Publishing, California Krug, S., 2006,

Kukulska-Hulme, A., & Pettit, J., 2007, Self-service Education: Smartphones as a catalyst for informal collective and individual learning, in Mlearn '07 mobile learning conference, 16-19th October 2007, Melbourne, Australia, available at: http://oro.open.ac.uk/16621/

Kukulska-Hulme, A., & Pettit, J., 2007a, Do smart devices make smart learners? in MLearn '07 mobile learning conference, 16-19th October 2007, Melbourne, Australia, pp.1-9, also available at: http://oro.open.ac.uk/16621/

Kuzel, A.J., 1992, Sampling in qualitative enquiry, In: Crabtree, B.F. & Miller, W.L. (Eds), 1992, Doing Qualitiative research, Sage, Newbury Park, California, pp31-44

Kvale, S., 1996, Interviews: An introduction to qualitative research interviewing. London: Sage, pp17-35

Kvavik, R.B., 2004, Convenience, Communications, and Control: How Students Use Technology EDUCAUSE, Center for Applied Research and University of Minnesota, Twin Cities, From: Kvavik, R.B., Caruso, J.B. & Morgan, G., 2004, ECAR Study of Students and Information Technology 2004: Convenience, Connection and Control, (Research Study Volume 5) Boulder, Colorado: ECUCAUSE Center for Applied Research, net.educause.edu/ir/library/pdf/

L'Abbé, K.A., Detsky, A.S., O'Rourke, K., 1987, Meta-analysis in clinical research, Annals of Internal Medicine 107 pp224-233

Laurillard, D., 2005, Presentation given at the HEA 2005 conference: Harnessing technology to improve the learning experience, Higher Education Academy.

Lawton, S., Montogomery, L., & Farmer, J., 2001, Survey and workshop initiative on community nurses' knowledge of the Internet, Comput Nurs 19(3), pp118-21

Learning Circuits Report, 2004, Available from: http://www.learningcircuits.org/mar2004/primer.html

Leeder, D., Wharrad, H.J., & Davies, T., 2002, Beyond Institutional Boundaries: RLO's for Multi-professional Education from: www.medgraphics.cam.ac.uk/ucel/docs/dublin...paper.doc

Lie, K.G. & Cano, V., 2001, Supporting diverse learners through a web site for teaching methods, Educational Technology & Society 4 (3)

Littlejohn, A. (Ed), 2003, Reusing online resources: a sustainable reusable to e-learning, London, Kogan Page

Lockitt, B., 2004, Adult, community and work-based learning: Cheadle, Uk, 3T productions

Longmire, W., 2000, A primer in learning objects, Learning Circuits, from: http://www.learningcircuits.org/mar2000/primer.html

Martin, D., 2010, Apple releases iPhone OS SDK 3.2 Beta 2 for iPad, *CNET,* from: http://reviews.cnet.com/8301-19512_7-10450636-233.html

Martinez, M., 2002, Designing learning objects to personalize learning, in: Wiley D.A., 2002, The instructional use of learning objects (chapter 1.1), Bloomington, Indiana, AIT & AECT, pp3-24

May, A.J., 2001, User-centred evaluation of complex IT&T within an operational environment, Journal of Research and Practice in Information Technology, 33 (1), pp83-97

Mayer, R.E., 2001, Multimedia Learning, Cambridge University Press, Cambridge

Mayer, R.E., Moreno, R., 1998, A split-attention effect in multimedia learning: evidence for dual processing systems in working memory, Journal of Educational Psychology, Vol.90, No.2, pp312-320

Mayer, R.E., Heiser, J., Lonn, S., 2001, Cognitive constraints on multimedia learning: when presenting more material results in less understanding, Journal of Educational Psychology, Vol.93, No.1, pp187-198

Mayer, R.E., Moreno, R., Boire, M., Vagge, S., 1999, Maximizing constructivist learning from multimedia communications by minimizing cognitive load, Journal of Educational Psychology, Vol .91, No.4, pp638-643.

McGee, P., 2003, Learning Objects: Bloom's Taxonomy and Deeper learning principles http://fle3.uiah.fi/

McGreal, R., 2004, (Ed.) Online Education using learning objects, London, Routledge/Farmer, pp13-14

Merkow, M. S., 2002, Learning Objects Spark an E-learning Revloution, Available from: www.techlearning.com/db_area/archives/WCE/mmerkov.html

Miller, J.,Shaw-Kokot, J.R., Arnold, M.S., Boggin, T., Crowell, K.E., Allegri, F. et al, 2005, A study of Personal Digital Assistants to enhance Undergraduate Nursing Education, *Journals of Nursing Education,* 44, pp.19-26

Mills, K., 2006, Discovering design possibilities through a pedagogy of multi-literacies, Journal of Learning Design 1(3), pp61-72, also available at: http://www.jld.qut.edu.au/

Moher, D. et al, 1999, Lancet 354, pp1896-1900, also available at: http://www.consort-statement.org/QUOROM.pdf.

Moher, D., Altman, D.G., Schulz, K.F. & Elbourne, D.R., 2004, (The CONSORT Group) Opportunities and challenges for improving the quality of reporting clinical research: CONSORT and beyond, Canadian Medical Association Journal 171 (4), p1503

Moreno, R., Mayer, R.E., 1999, Cognitive principles of multimedia learning: The role of modality and contiguity, Journal of Educational Psychology, Vol.91, No.2, pp358-368

Morse, J.M., 1995, The significance of saturation, Qualitative Health Research 5 (2), pp147-149

Mulrow, C.D., 1994, Systematic Reviews: Rationale for systematic reviews (3 September) Education and debate, BMJ 309, pp597-599

Muzio, J., Heins T. & Mundell, R., 2001, Experiences with Reusable eLearning objects: from theory to practice, from: http://www.cedarlearning.com/CL/elo/elearningObjects_sml.pdf

Myers, S. T & Hasse, J.E. 1989 Guidelines for integration of quantitative and qualitative approaches Nursing Research 38 pp299-301

Nah, K.C., White, P., & Sussex, R. (2008) The potential for using a mobile phone to access the internet for learning EFL listening skills within a Korean context, *ReCALL*, 20, pp.331-347

Naidu, S., 2002, Designing and evaluating instruction for e-learning, in: Rogers, P. (Ed), 2002, Designing instruction for technology-enhanced learning (Chapter VIII), Hersey, PA, Idea Group Publication, pp134-159

Nash, S. S., 2005, Learning objects, learning object repositories, and learning theory: Preliminary best practices for online courses, Interdisciplinary Journal of Knowledge and Learning Objects 1, pp217-228, Available from: http://ijklo.org/Volume1/v1p217-228Nash.pdf

NCIHE, 1997, Higher Education in the Learning Society: The Dearing Report of Inquiry into Higher Education, National Committee, London, HMSO

Nesbit, J., Belfer, K. & Vargo, J., 2002, A convergent participant model for evaluation of learning objects, Canadian Journal of Learning and Technology, 28 (3)

Neumeier, P., 2005, A closer look at blended learning – parameters for a blended learning environment for language teaching and learning, ReCALL 17, pp163-178

New London Group, 2000, A pedagogy of multiliteracies: Designing social futures, In Cope, B. & Kalantzis, M. (Eds.), Multiliteracies: Literacy learning and the design of social futures
Melbourne: Macmillan, pp9–38

Multi-literacy pedagogy, from: Mills, K., 2006, Discovering design possibilities through a pedagogy of multi-literacies, Journal of Learning Design 1(3), pp61-72, also available at: http://www.jld.qut.edu.au/

Newman M., Papadopoulos I., Sigsworth J., 1998, Barriers to Evidence-based Practice, Clinical Effectiveness in Nursing 2, pp11-20

Nielsen, J., & Landauer, T.K., 2000, March Alertbox column, Why You Only Need to Test with 5 Users, from: http://www.useit.com p148; and Nielsen, J., & Landauer, T.K., 2000a, A mathematical model of the finding of usability problems, In CHI '93: Proceedings of the SIGCHI conference on Human factors in computing systems, pp206–213, New York, NY, USA, 1993, ACM Press

Nusca, A., 2009, Smartphone vs. feature phone arms race heats up: which did you buy? ZDNet, from: http://www.zdnet.com/blog/gadgetreviews/smartphone-vs-feature-phone-arms-race-heats-up-which-did-you-buy/6836 retrieved 19 May 2010.

Oliver, M. et al, 2000, Research Study on the effectiveness of resources, tools and support services used by practitioners in designing and delivering e-learning activities, Acets 3

Olivier, B. & Liber, O., 2003, Learning content interoperability Standards in Littlejohn A (Ed), 2003, Reusing online resources: a sustainable approach, London, Kogan Page

Orrill, C.H., 2002, Learning objects to support inquiry-based on-line learning, in: Wiley, D.(Ed), The instructional use of learning objects (Chapter 2.3), Bloomington, Indiana, AIT & AECT, pp131-148

Oxman, A.D., & Guyatt, G.H., 1992, A consumer's guide to subgroup analyses, Annals of Internal Medicine 116, pp78-84

Ottenbacher, K.J., 1995, Why rehabilitation research does not work as well as we think it should, Archives of Physical, Medical Rehabilitation, 76, pp123-9

Pachler, N., Bachmair, B., & Cook, J., 2010, Mobile learning: Structures Agency Practices, Publisher & Place (Introduction pp1-2, chapter 2 pp59-71 & 155-174, chapter 11 pp273-296)

Paczkowski. J., 2010, *Initial iPad demand greater than initial iPhone demand,* All things digital, from: http://digitaldaily.allthingsd.com/20100223/initial-ipad-demand-greater-than-iphone-demand/

Parahoo, K., 2009, Grounded theory: what's the point?, Nurse Researcher, Vol.17, No.1, p9

PC Magazine, 2010, Smartphone definition from PC Magazine Encyclopaedia, from: http://pcmag.com/encyclopedia_term.php?qid=310

Peters, 2007, M-Learning: Positioning educators for a mobile connected future, The International Review of Research in Open and Distance Learning, Vol.8, No.2, pp492-3831

Petry, Mouton & Reigluth, 1987, Cited from Newman M., Papadopoulos I., Sigsworth J., 1998, Barriers to Evidence-based Practice, Clinical Effectiveness in Nursing 2, pp11-20

Phillips, E., & Pugh, D. (2000). How to get a PhD (2nd ed.). Buckingham: Open University Press.

Polit, D.F. & Hungler, B.P., 1999, Nursing Research: Principles and methods Philadelphia, J.B. Lippincott Company

Postema, T. & Markham, S., 2002, Web-based learning environments: developing a framework for evaluation, from: Sheard, J. & Markham, S., 2005), Assessment and Evaluation in Higher Education, Vol. 30, No. 4, pp353-368

Powell, S. 2005, Ecology of E-learning Philosophies, Retrieved 1-12-05 Article no longer available online. Referred to in: Bradshaw, P., Powell, S. And Terrell, I., 2005, Devoloping Engagement in Ultralab's Online Communities of Enquiry, Innovations in Education and Teaching International, Vol.42, No.3, pp205-215

Prosser, M., 2005, Academy Exchange, Higher Education Academy, Issue 1 Summer, p8-9

Prosser. M. & Trigwell, 1997, Design patterns for complex learning, Journal of Learning Design, 1(3), pp82-91, also available at: http://www.jld.qut.edu.au/, and on the 'Complexity and Education' website at: http://www.complexityandeducation.ualberta.ca

QUOROM, 2000, Using this template: Systematic reviews and meta-analysis, from http://www.consort-statement.org/QUOROM.pdf)

Ramsden P., 1992, Learning to teach in Higher Education, London Routledge

Riding & Rayner, 1999, Towards A Comprehensive Learning Object Metadata, In Yahya, Y. & Yusoff, M., 2008, Incorporation of Context to Stipulate Meaningful Learning and Enhance Learning objects, Interdisciplinary Journal of E-Learning and Learning Objects, Volume 4, pp3-24

Robinson, J.M., Frey, J.G., Stanford-Clark, A.J., Reynolds, A.D., & Bedi, B.V., 2010, Chemistry by mobile phone (or how to justify more time at the bar), from: http://eprints.soton.ac.uk/28632/01/481.pdf

Rogers, P., 2002, Designing instruction for technology- Enhanced learning, Hersey, PA, Idea Group Publishing

Robertson, M. & Fluck, A., 2004, What makes a good learning object? AARE pp1-12

Rohse, S., & Anderson, T., 2006, Design patterns for complex learning, Journal of Learning Design, 1(3), pp82-91, also available at: http://www.jld.qut.edu.au/, and on the 'Complexity and Education' website at: http://www.complexityandeducation.ualberta.ca

Romero, M. & Wareham, J., 2009, Just in time mobile learning model based on context awareness information, January to April 2009, IEEE Computer Society, *Learning Technology Newsletter*, Vol.11, Issue 1 & 2 pp.4-6

Romiszowski, A.J.,1992, The Selection and Use of Instructional Media, Second edition, Kogan Page, London

Rosenthal, R. & DiMatteo, M. R., 2001, Meta-analysis: Recent developments, in: Quantitive methods, Annual Review of Psychology, Vol. 52, pp59-82

Rossi, P.H., Freeman, H.E. & Lipsey, M.W., 1999, Evaluation: a systematic approach, Sage Publications, London

Ryan, P., & Bernard, A, 2003, Field Methods, Sage Publications 15(1)m pp85–109

Sackett, D.L., Rosenburg, W.M.C., Gray, J.A.M., Haynes, R.B. & Richardson, W.S., 1996, Evidence-based medicine: what it is and what it isn't (Editorial), BMJ 312, pp71

Sanger, F., Nicklen, S., & Coulson, A.R., 1977, DNA sequencing with chain-terminating inhibitors, Proc Natl Acad Sci U S A 74: pp5463–5467

Schmeck, Ribich, and Ramanaiah (1977) cited in Riding & Rayner, 1999 Cited in Yusoff 2008 Towards A Comprehensive Learning Object Metadata: Incorporation of Context to Stipulate Meaningful Learning and Enhance Learning Objects, Vol.4, Interdisciplinary Journal of E-Learning and Learning Objects (Formerly the Interdisciplinary Journal of Knowledge and Learning Objects)

Scriven, M. 1980, Cited in Sheard, J & Markham, S., 2005, Web-based learning environments: developing a framework for evaluation, Assessment and Evaluation in Higher Education, Vol. 30 No. 4, pp353-368

Sheard, J. & Markham, S., 2005, Web-based learning environments: developing a framework for evaluation, Assessment and Evaluation in Higher Education, Vol. 30, No. 4, pp353-368

Shulman, L., 2002, Making differences: A table of learning, *Change*, 34, (6), also available at: http://www.carnegiefoundation.org/elibrary/docs/making_differences.htm

Sim, J. & Reid, 1999, Professional perspectives statistical inference by confidence intervals: Issues of interpretation and utilisation, Physical Therapy, Vol.79, No.2, pp186-195

Sims, R., 2006, Beyond instructional design: Making learning design a reality, Journal of Learning Design, Vol.1, No.2, pp1-7 Also available at: http://www.jld.qut.edu.au

Schmidt, H. G., 1993, Foundations of problem-based learning: some explanatory notes, *Medical Education* 27, pp422-432

SCIE, 2006, The Conduct of Systematic Research Reviews for SCIE Knowledge reviews, Knowledge review booklet guidelines

Song, F. 1999, Exploring heterogeneity in meta-analysis: Is the L'Abbé plot useful? Journal of Clinical Epidemiology, 52(8), pp725-730

Spector, T.D., & Thompson, S.G., 1991, The potential and limitations of meta-analysis, Journal of Epidemiology and Community Health 45, pp89-92

Sproull, L., & Kiesler, S., 1992, Connections, Cambridge , MA : MIT Press, Also: The University of Illinois Online (2003), Available from: http://www.online.uillinois.edu

StatsDirect , 2010, http://www.statsdirect.com/help/meta_analysis/heterogeneity_in_meta_analysis.htm

Stern, J., 1980, Grounded theory methodology: its use and processes, Image, 12 (1), pp20-23

Stewart, L.A., & Parmar, M.K.B., 1993, Meta-analysis of the literature or of individual patient data: is there a difference? Lancet 341, pp418-422

Stockwell, G., 2008, Investigating learner preparedness for and usage patterns of mobile learning, *ReCALL* 20, Cambridge University Press, pp253-270

Strauss, A.L., 1987, Qualitative analysis for social scientists, Cambridge University Press, NY, p107

Strauss, A.L. & Corbin, T.1990, cited in: Holloway, I. & Wheeler, S., 1996, Qualitative Research for Nurses, Blackwell Science, Berlin, Germany, p106

Thacker, S.B., 1990, Meta-analysis: a quantitative approach to research integration (Chapter 8) in: Gelijns, A., Modern Methods of Clinical Investigation, Instititute of Medicine, Medical Innovation at the crossroads, Vol.1, pp89-91

Thomason, L., 2004, Design Tip: Web Site Usability Checklist, MetMechanic, Inc. Feb. 2004 (Part 2), Vol. 7, No. 4, also available at: http://www.netmechanic.com/news/vol7/design_no4.htm

Thompson, S.G., 1994, Systematic Review: Why sources of heterogeneity in meta-analysis should be investigated (Education and debate), BMJ 309, pp1351-5

Thompson, S.G., Smith, T.C., Sharp, S.J., 1997, Investigating underlying risk as a source of heterogeneity in meta-analysis, Statistics in Medicine 16, pp2741-58

Thompson, S.G. & Higgins, J.P.T., 2002, How should meta-regression analyses be undertaken and interpreted? Statistics in Medicine 21, pp1559-1573

Thorpe, M., Kubiak, C. & Thorpe, K., 2003, Designing for reuse and versioning, in: Littlejohn, A. (Ed), Reusing online resources: a sustainable approach to e-Learning (Chapter 9), London, Kogan Page, pp106-118

TLT Group, 2010, available at www.tltgroup.org

Topolsky, J., 2010 Apple iPad: The definitive guide (so far), Engadget, available from: http://engadget.com/2010/01/29/apple-ipad-the-definitive-guide-so-far/

Traxler, J., 2009, The Current State of Mobile Learning, pp9-25 - chapter in Ally, M., 2009, (Ed.) Mobile Learning: Transforming the Delivery of Education, Athabasca University Press (Original publication: Traxler, J. 2007, Defining, Discussing and Evaluating Mobile Learning: the moving finger writes, The International Review of Research in Open and Distance Learning, Vol 8, No 2, p19

U.K.C.C., 1999, Fitness for Practice, London, UKCC

Van Der Vleuten C., Dolmans D. Scherpbier A., 2000, The need for evidence in education, Medical Teacher, 22(3), pp246-250

Vernon D.T.A., & Blake R.L., 1993, Does PBL Learning Work? A Meta-analysis of Evaluative Research, Academic Medicine 69 (7), pp550-563

Verstraete, M., 2002, Value and limitation of meta-analysis, Pathoph. Haemost Thromb, 32, pp278-28

Waddell, D.L., 1991, The Effects of Continuing Education on Nursing Practice: a Meta-analysis, Journal of Continuing Nursing Education 22, pp113-118

Wang, S., 2008, Ontology of Learning Objects Repository for Pedagogical Knowledge Sharing, Interdisciplinary Journal of E-Learning and Learning Objects, Vol.4, pp4-10 (formerly the Interdisciplinary Journal of Knowledge and Learning Objects, *University of Massachusetts Dartmouth, MA, USA)*

Waycott, J., 2002, Reading with new tools: an evaluation of personal digital assistants as tools for reading course materials, *ALT-J*, Vol.10, No.2, pp38-50

Weigel, V., 2002, Deep Learning for a digital age: Technologies untapped potential to enrich higher education, New York, Jossey Bass

Weller, M., Pegler, C., & Mason, R., 2003, Putting the pieces together: What working with learning objects means for the educator, cited in Christiansen, J. & Anderson, T., 2004, Feasibility of course development based on learning objects: research analysis of three case studies, available from: (second page) http://www.itdl.org/Journal/Mar_04/article02.htm,

West, J., Tuffnell, D. & Kleijnen, J., 2002, Promoting evidence: systematic reviews (What they are, how they are carried out, and how midwives can access them on the Internet British Journal of Midwifery, 10(12), pp771-5

Whalley. P., 1996, An alternative rhetoric for hypertext, in Goodfellow, R. & Kukulska-Hulme, A., 1996, Evaluating new technology in learning and teaching: A summary of recent research at the Institute of Educational Technology CITE Report 225

White, A., Allen, P., Goodwin, L., Breckinridge, D., Dowell, J., & *Garvy, R., 2005, Infusing PDA technology into nursing* education. Nurse Educator, 30, pp150-154

Wiley, D.A., 1999a, The post-LEGO learning object, from: http://wiley.ed.usu.edu/docs/post-lego.pdf

Wiley, D.A., 1999b, So what do I do with learning object? From: http://wiley.ed.usu.edu/docs/instruct-arch.pdf

Wiley, D.A., 2000, Learning object design and sequencing theory (Doctoral Dissertation), Brigham Young University

Wiley, D.A., 2002a, CAREER grant – A mediated action Study of learning object use in online Learning communities from http://wiley.ed.usu.edu/docs/career.pdf

Wiley, D.A., 2002b, Connecting learning objects to instructional design theory: a definition, a metaphor, and a taxonomy, from: Wiley, D.A. (Ed), 2002c, The instructional use of learning objects (chapter 1.1), Indiana Bloomington, Agency for Instructional Technology and Association for Educational Communications and Technology, pp3-24

Wiley, D. A., 2002c, The instructional use of learning objects, Bloomington Indiana, AIT/AECT, pp3-24

Wiley, D.A., 2002d, Learning objects – a definition, in: Kovalchick, A. & Dawson, K. (Eds) 2002, Educational technology: an Encyclopaedia, Santa Barbara, ABC-CLIO

Wiley, D.A., 2002e, Learning objects need instructional Design theory in Rossett, A. (Ed), 2001/2002 ASTD Distance learning Yearbook, New York, McGraw Hill

Wiley, D.A., 2002f, A unified design framework for Learning objects and educational Discourse from: http://wiley.ed.usu.edu/docs/lo_do.pdf

Wiley, D. A. (2003). Learning objects: Difficulties and opportunities. *Academic ADL Co-Lab News Report: No. 152-030406.* [Online]. Retrieved April 7, 2003, from http://wiley.ed.usu.edu/docs/lo_do.pdf.

Wiley, D.A., Recker, M.M. & Gibbons, A., 2000a, Getting axiomatic about learning objects from: http://reusability.org/granularity.pdf

Wiley, D.A. Recker, M.M., & Gibbons, A., 2000b, A reformulation of object granularity, available from: http://reusability.org/granularity.pdf

Wilkinson, S., Dennick, R.,& Rae, H., 2006, A Touchstone for online summative assessment, Issues and news on learning and teaching in medicine, dentistry and veterinary medicine (summer 2006), The Higher Education Academy, pp32-3

Williams, D.D., 2002, Evaluation of learning objects and Instruction using learning objects In Wiley, D. (Ed) The instructional use of learning objects (chapter 3), Bloomington, Indiana, Agency for Instructional Technology/Association for Educational Communications & Technology, pp173-200

Winters, J., Cotton, D., Gavin, J., & Yorke, J.D., 2010, Effective E-learning? Multi-tasking, distractions and boundary management by graduate students in an online environment, ALT-J, Research in Learning Technology, Vol. 18, No.1, pp71-83

Wishart, J., 2006, Personal Digital Assistants – teachers prefer the personal, paper given at *Mlearn'06, 5th world conference on mobile learning*, Banff, October 22-26th 2006, Canada, p15

Wishart, J., 2008, Challenges faced by Modern foreign language teacher trainees in using handheld pocket PCs (PDAs) to support their teaching and learning, Cambridge University Press, *ReCALL*, 20, pp.348-360

Wolcott, H.F., 1994, Transforming qualitative data: Description, analysis and interpretation, Sage, Thousand Oaks, California, p94

Wolf, F., 2000, Lessons to be learned from evidence-based medicine: practice and promise of evidence based medicine and evidence-based education, Medical Teacher 22(3), pp251-5

Wood, I., 1998, The effects of Continuing Professional Education on the Clinical Practice of Nurses: a Review of the Literature, International Journal of Nursing Studies 35, pp125-131

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L., 1997, Program evaluation: Alternative approaches New York, Longman and practical guidelines (2nd Ed)

Yahya, Y. & Yusoff, M. 2008, Towards A Comprehensive Learning Object Metadata: Incorporation of Context to Stipulate Meaningful Learning and Enhance Learning Object Reusability, Interdisciplinary Journal of E-Learning and Learning Objects Volume 4, (Formerly the Interdisciplinary Journal of Knowledge and Learning Objects), Vol.4, pp335-371

Yen, N.Y. Shih, T.K. Chao, L.R. Qun. J., 2010, Ranking Metrics and Search Guidance for Learning Object Repository, Learning Technologies, IEEE Transactions (July-Sept), Vol.3, Issue 3, pp250-264

Zetterberg, H.L., 1963, On theory and verification in sociology, Totowa, New Jersey, Bedminster Press, pp52-56

12. BIBLIOGRAPHY

Arrington, M., 2010, The unauthorized TechCrunch iPad review, *TechCrunch*, from: http://techcrunch.com/2010/04/02/the-unauthorized-techcrunch-review/

Bandura, A., 1986, Social Foundations of Thought and Action, London, Prentice Hall

Barritt, C., & Alderman Jr., F. L., 2004, Creating a reusable learning objects strategy leveraging information and learning in a knowledge economy. Pfeiffer

Barritt, C., Lewis, C. & Weiseler, C., 1999, Cisco Systems Reusable Information - Object Strategy: Definition, Creation Overview and Guidelines, from: http://www.cisco.com/warp/public/779/ibs/solutions /learning/whitepapers/el_cisco_rio.pdf

Barritt, C., & Lewis, D., 2001, Reusable learning object strategy, from http://www.cisco.com/warp /public/10/wwtraining/elearning/learn/whitepaper_docs/rlo_strategy_v3-1.pdf

Baumgart, D.C., 2005, Personal Digital Assistants in Health Care: experienced clinicians in the palm of your hand? *Lancet*, 366, pp.1210-1222

Buchwald, H., Avidor, Y., Braunwald, E., Jensen, M.D., Pories, W., Fahrbach, K., & Schoelles, K., 2004, Bariatric Surgery: A Systematic Review and Meta-analysis (October 13), JAMA. 292, Vol. 292 No.14, pp1724-1737

Campbell, L.M., 2003, Engaging with the learning object economy In: Littlejohn A (Ed) Reusing online resources: a sustainable approach London Kogan Page

Clarke, M., 2004, Systematic reviews and the Cochrane Collaboration, Available at: www.cochrane.org/docs/whycc.htm).

Clough, G., Jones, A.C., McAndrew, P., & Scanlon, E., 2007, Informal learning with PDAs and Smartphones, Journal of Computer assisted learning, 24 (5), pp.359-371

Conole, G., 2004, Research Study on the Effectiveness of Resources, Tools and Support Services used Resources, Tools and Support Services used by Practitioners in Designing and Delivering E-Learning Activities' (Report on the effectiveness of tools for e-learning, Resources, Tools and Support Services used)

Crook, C., 2002, The campus experience of networked learning. In C. Steeples & C. Jones (Editors), Networked Learning: perspectives and issues, London, Springer, pp293–308

Davis, D., Thomson, M.A., O'Brien, M.A., Freemantle, N., Wolf, F.M., Mazmanian, P., & Taylor-Vaisey, A., 1999, Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? JAMA 282 (9), pp867-874

Davies, P., 2000, Approaches to evidence based teaching, Medical Teacher 22(1), pp14-21

Denscombe, M., 1998, The Good Research Guide: For Small-scale Social Research Projects, Buckingham: Open University Press

Easterbrook, P.J., 1991, Berlin, J.A., Gopalan, R., & Matthews, D.R., Publication bias in clinical research, Lancet, 337, pp867-872

Egger, M., Davey-Smith, G., Schneider, M., & Minder, C., 1997, Bias in meta-analysis detected by a simple graphical test (13 September), BMJ 315, pp629-634

Egger, M., Davey-Smith, G. & Phillips, A.N., 1997, Meta-analysis: Prinicples and procedures (Education and Debate – 6th December), BMJ 315, pp1533-1537

Foreman D. & Nyatanga L., 1999, The evolution of shared learning: some political and professional imperatives, Medical Teacher 21, p489-96

Hara, N., & Kling, R., 2000, Student distress in a web-based distance education course, Information, Communication & Society 3(4), pp557–579

Horsburg, M., Landin, R. & Williamson, E ., 2001, Multiprofessional learning: the attitudes of medical, nursing and pharmacy students to learning, Medical Education 35, pp876-83

Hughes, M.D., Freedman, L.S., & Pocock, S.J., 1992, The impact of stopping rules on heterogeneity of results in overviews of clinical trials, Biometrics 48, pp41-53

Human Factors, 2009. Available from: http://www.humanfactors.com/training/usabilitytesting.asp

ISDL, 2004, Instructional Design Theory, International Skills Development Ltd, available from: http://www.interskills.com/E-design.htm

Jadad, A., 1999, Promoting partnerships: challenges for the Internet age, BMJ 319, pp761–771

Jadad, A.R., Haynes,B., Hunt, D. & Browman, G.P., 2000, The Internet and evidence-based decision-making a needed synergy for efficient knowledge management in health care, CMAJ 162, pp362–364

Jadad, A., Sigouin, C., Cocking, L., Booker, L., Whelan, T. & Browman, G., 2001, Internet use among physicians, nurses and their patients, JAMA 286, pp1451–1452

James, G., 2002, Advantages and disadvantages of online Learning, from: http://www.comminit.com/en/node/210058/307

Justham, D., & Timmons, S., 2005, An evaluation of using a web-based statistics test to teach statistics to post- registration nursing students, Nurse Education Today, 25(2), pp156-63

Kennedy, 2004, from Sims, R. (2006) Beyond instructional design: Making learning design a reality, Journal of Learning Design 1(2), pp1-7, also available at http://www.jld.qut.edu.au

Kruse, K., 2002, Using the web for learning: advanatges and Disadvantages, available from: http://www.e-learningguru.com/articles/art1_9.htm

Kukulsha-Hulme, A.& Petit, J., 2008, Practitioners as educators: Emergent practice in Personal mobile teaching, Learning, work and leisure, pp135-155

Leung, W., 2002, The use of the Internet and information technology to facilitate teaching evidence base practice: a case study, Nurse Education in Practice, 2(3), pp181-9

Maguire, P., & Pitceathly, C., 2002, Key communication skills and how to acquire them, BMJ 325, pp697-700, from: http://bmj.com/cgi/content/full/325/7366/697

Microsoft Website, 2010, Available from http://www.microsoft.com/windosmobile/en-gb/devices/smartphones.mspx?13=&14=&31.35=1&24=1&56=1&25=1

Morison, S., Boohan, M .,Jenkins, J., & Moultray, M., 2003, Facilitating undergraduate inter-professional learning in healthcare:comparing classroom and clinical learning for nursing and medical students learning, Health and Social Care, 2(2) pp92-104

Muller, S., 1984, 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 (Nov) Part 2

Nah, K.C., White, P., & Sussex, R., 2008, The potential for using a mobile phone to access the internet for learning EFL listening skills within a Korean context, *ReCALL* 20, pp331-347

Norman, G. & Schmidt, H., 1992, The psychological basis of PBL learning: A review of the evidence, Academic Medicine 67(9), pp557-564

O'Malley, C., Vavoula, G., Glew, J.P., Taylor, J., Sharples, M. & Lefrere, P., 2003, Guidelines for learning/teaching/tutoring in a mobile environment, available at: www.mobilearn.org/results/results.htm

Petrosino A., Boruch R., Rounding, C., McDonald, S., Chalmers, I., 2000, The Campbell Collaboration Social, Evaluation & Psychological, Educational and Crim- inological Trials Register (C2-SPECTR) to facilitate the preparation and maintenance of systematic reviews of social and educational interventions, Research in Education, 14 (3&4), pp206-217

Prensky, M., 2001, Digital Natives Digital Immigrants, On The Horizon, NCB University Press, 9(5), Retrieved March 2007. This is available from: http://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf

Russell, R., 1999, Experience-Based Learning Theories, The Informal Learning review, http://www.informallearning.com/archive/1999-0304-a.htm

Schraefel, M.C., Andre, P., & Van Kllek, M., 2010, This time it's personal: from PIM to the perfect digital assistant, pp.1-4, from: http://eprints.lancs.ac.uk/22928/1/getPDF.pdf

Schofield, I., 2002, (Feb.) Log on to develop practice (Web-based gerontological nursing demonstration project to facilitate evidence based practice, developed by Scotland's Nursing and Midwifery Practice Development Unit, Nursing Times, 98(7), pp38-9

Seale, J., 2006, (Ed), Individual Enthusiasm to Institutional Implementation: A Review of Learning Technology in Post Compulsory Education, Swets & Zeitlinger, New Zealand.

Shultz, K.F., Chalmers, I., Hayes, R.J., & Altman, D.G.,2010, Empirical evidence of bias: dimensions of methodological quality are associated with estimates of treatment effects in 250 controlled trials form 33 meta-analyses, in press JAMA.

Sharpe, R., Benfield, G., Lessner, E., & DeCicco, E., 2005, Scoping study for the pedagogy strand of the JISC e-learning programme,from: http://www.jisc.ac.uk/elp_learneroutcomes.html

Selwyn, N., 2005, Web 2.0 applications as alternative environments for informal learning - a critical review, Paper for OECD-KERIS expert meeting - Session 6 – Alternative learning environments in practice: using ICT to change impact and outcomes, Institute of Education, University of London, UK

Varlamis, & Apostalakis, 2006, The OSEL Taxonomy for the Classification of Learning Objects, Interdisciplinary Journal of Knowledge and Learning Objects, Vol.2,

Vrasidas, C & McIssac, M., 1999, Principles of Pedagogy and evaluation of web-based learning, Oral paper given at ICEM International Conference for Educational Media 37:2 Ljubljana, 2000, Available at: www.seamonkey.ed.asu.edu/~mcisaac/ICEM99/presentationppt/sld008.htm ,Arizona State University. Also available at: http://www.tandf.co.uk/journals

Waycott, J. & Kukulsha-Hulme, A., 2003, Students' experiences with PDAs for reading course materials, *Personal and ubiquitous computing* 7 (1), pp30-43

W.H.O., 1993, Increasing the relevance of Education for Health professionals. Report of a WHO study group on Problem Solving Education for Health professionals Technical Report Series No. 838, Geneva, WHO

Wikipedia, 2010, from http://en.wikipedia.org/wiki/IPhone

Wikipedia, 2010, from http://wikipedia.org/wiki/PDA

Woodward C., 1997, What can we learn from programme evaluation studies in medical education, In: Boud, D., & Feletti,G. (Eds), The Challenge of Problem Based Learning (2nd ed.), London, Kogan Page, pp294-307

13. PUBLICATIONS FROM THIS THESIS TO DATE

Calbraith, D. 2006, Systematic approaches to e-learning methodology and evaluation, Oral paper given at the International iPED conference, October, Coventry University, Published in Conference proceedings.

Calbraith, D. & Dennick, R., 2009a, Pedagogical Strategies for Learning Objects – Published Abstract, IATED publishing. Oral paper also given at the ICERI International conference, Madrid, Spain, 16-18th November 2009, Also available at www.iated.org/iceri2009/index.php?section=19

Calbraith, D. & Dennick, R., 2009b, Pedagogical strategies for learning objects – Published full paper, IATED publishing. Oral paper also given at the ICERI International conference, Madrid, Spain, 16-18th November 2009, Also available at www.iated.org/iceri2009/index.php?section=19

Calbraith, D., 2010a, Discovering a reliable quality base for E-learning evaluation, Abstract and oral paper published in the RCN 2009 International Research Conference proceedings, Blackpool

Calbraith, D., 2010b, Analysing Student experiences using Touchstone, Abstract and oral paper published in the RCN 2009 International Research Conference proceedings, Blackpool

Calbraith, D. & Dennick, R., 2011, Producing Generic Principles and Pedagogies for Mobile Learning: A Rigorous Five Part model, Chapter 2 in: Kitchenham, A. (Ed), 2011, Models for Interdisciplinary Mobile Learning: Delivering Information to Students, IGI Global, Pennsylvania (In press).

13. APPENDICES

APPENDIX 1 – *Generic Student Survey Questions*

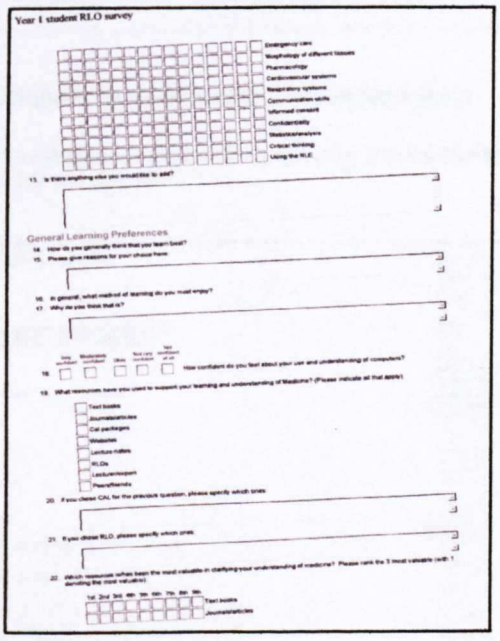
What specific areas in your module do you find most enjoyable?
What specific ways of learning/approaches do you most enjoy?
What specific areas in your module do you find most difficult?
What specific ways of learning/approaches do you find most difficult?
What specific area do you think an RLO would be most beneficial?
In which areas of your curriculum would it be MOST useful to have a high quality RLO?
Do you have regular access to the internet?
Are you confident about your computer skills?

APPENDIX 2 – Touchstone Student Questionnaire: What The Survey Writer Sees

This example shows a section of the survey during the creation stage. This is what the page looks like when entering question choices in Touchstone:

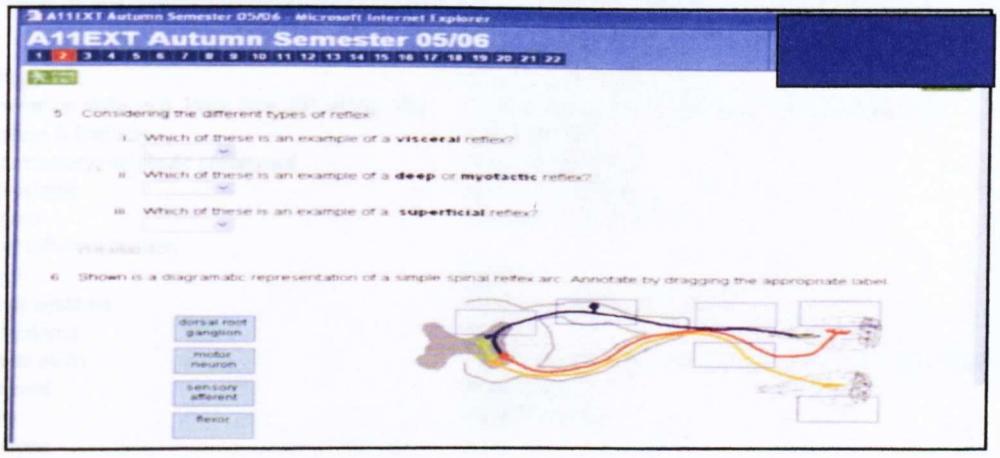
	Question	Type	Modified	Order
13	Is there anything else you would like to add?	Text Box	21/11/06	
14	How do you generally think you learn best?	Multiple choice	21/11/06	
15	Please give reasons for your choice here	Text Box	21/11/06	
16	In general, what method of learning do you most enjoy?	Multiple choice	21/11/06	
17	Why do you think that is?	Text Box	21/11/06	
18	How confident do you feel about your use and understanding of computers?	Likert scale	21/11/06	
19	What resources have you used to support your learning, and understanding of medicine?	Multiple response	21/11/06	
20	If you chose CAL for the previous question, please specify which ones:	Text Box	21/11/06	
21	If you chose RLO, please specify which one	Text Box	21/11/06	
22	Which resources are/has been the most valuable in clarifying your understanding of medicine? Please rank the 3 most valuable (with 1 denoting the highest and 5 the lowest)	Ranking	21/11/06	

APPENDIX 3 – Touchstone Student Survey Questionnaire



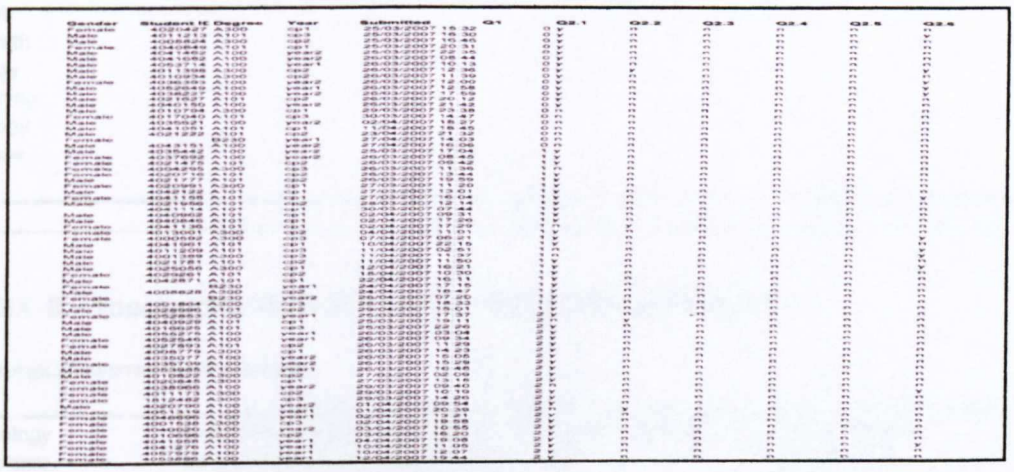
APPENDIX 4 – Touchstone Student Questionnaire: What The Student Sees

This shows an example of a survey section after saving. What the student sees is very different to the creation pages:



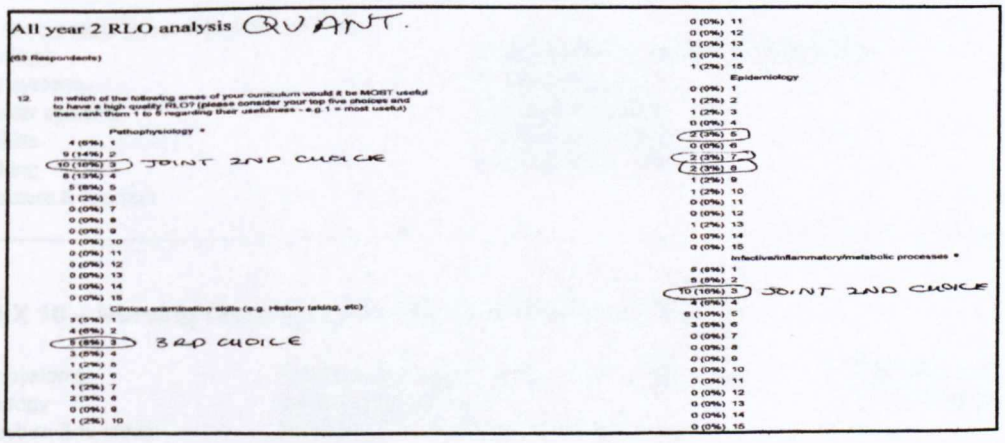
APPENDIX 5 – Touchstone Student Questionnaire: Raw Data

The survey's raw data is collected by Touchstone and saved as an excel spreadsheet. Files and can be printed out with or without grid as required. The survey's raw data is shown below:



APPENDIX 6 – Touchstone Student Questionnaire – Collated data

A collated data option is available in Touchstone which breaks down data into percentages for each year group. This made it easy to pick out top student choices for LO development:



APPENDIX 7 – Touchstone Student Questionnaire - Year by Year Breakdown

A year by year breakdown of the most needed subjects are given for B.Med.Sci students below:

Year 1	Year 2
Pathophysiology	*All of Year 1 subjects plus:*
Clinical examination skills (e.g. Peak flow, BP taking, etc)	Clinical examination skills (e.g. History taking, etc)
Normal Structure & function	Epidemiology
Infective/inflammatory/metabolic processes	Fetal development
Respiratory systems	Alimentary system
Emergency care	Urinary/kidney systems
Morphology of different tissues	
Pharmacology	**Year 3**
Cardiovascular systems	*All of Year 1& 2 subjects plus:*
Respiratory systems	Clinical:
Communication skills	Clinical Practice, Morphology of different tissues, Therapeutics
Informed consent	Non-clinical:
Confidentiality	Research skills
Statistics/analysis	
Critical thinking	
Using the NLE	
Year 4	**Year 5**
All of years 1-3 plus:	*As Year 4*
Clinical:	
Child Health	
Elderly Health	
Dermatology	
Otorhinolaryng.	
Opthamlology	
Obs & Gynae	
Psychiatry	

APPENDIX 8 – Touchstone Student Questionnaire - Common Subjects

Therefore subjects common to all years are:

Pathophysiology	Clinical examination skills (e.g. Peak flow, BP taking, etc)	Informed consent
Emergency care	Infective/inflammatory/metabolic processes	Using the NLE
Respiratory systems	Normal Structure & function	Critical thinking
Pharmacology	Morphology of different tissues	Confidentiality
Communication skills	Cardiovascular systems	Statistics/analysis

---

APPENDIX 9 – Nursing Questionnaire - Most Needed Subjects

Using new clinical skills (most enjoyable subject)	
Pathophysiology	Infective/inflammatory/metabolic processes
Respiratory systems	Emergency care
Cardiovascular systems	Respiratory systems
Research skills	Literature searching
Critical thinking	Detecting deterioration
Normal Structure & function	

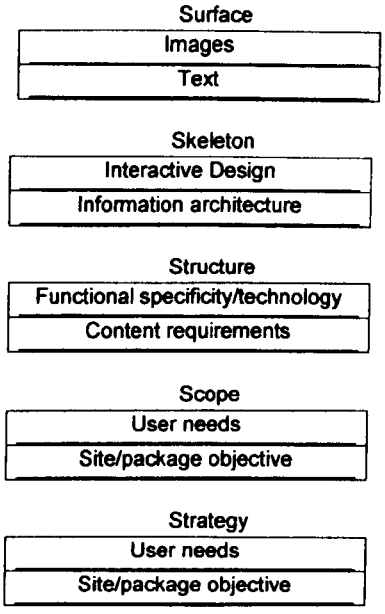
APPENDIX 10 – Nursing Questionnaire - Subjects Common To All

Respiratory systems	Clinical examination skills (e.g. History taking, etc)	Emergency care
Pathophysiology	Detecting deterioration	Cardiovascular systems
Normal Structure & function	Infective/inflammatory/metabolic processes	

APPENDIX 11 – Garrett's (2003, p23) Five 'planes'

Garrett's five planes are:

i) surface plane – images and text;
ii) skeleton plane – placement of buttons/tabs, photos, text, etc to optimise effect and efficiency;
iii) structure plane – navigation;
iv) scope plane – features and functions; and
v) strategy plane – developers/users explicit objectives for the media



APPENDIX 12 – Krug's 'Lost our lease' Usability Testing

An example of 'Lost-our-lease' usability testing (Krug 2006, p152) compared with 'tradtional-type' testing is as follows:

	TRADITIONAL TESTING	LOST-OUR-LEASE TESTING
No of users per test	Usually 8 or more to justify the set up costs	3 or 4
Recruiting	Select people to match the target audience	Anybody who uses the web will do
Where to test	Usability lab, one-way mirror observation	Any office/room
Who does the testing	An experienced usability professional	Any reasonably patient human being
Advance planning	Tests are scheduled weeks in advance, labs are booked and participants recruited	Anywhere with little advance planning
Preparation	Drfat, discuss & revise a test protocol	Decide what you're going to show
What/when do you test?	If no large budget, test once when almost complete	Run small tests throughout the development process
Cost	£2,500-£7,500 plus	£150 (£25-£50 per user)
What happens afterwards	Written report. Development team discuss what will be changed and how	The observer writes notes. Material can be modified the same day

Lost-our-lease testing was used throughout this thesis except in the following: i) Advance planning was required for the Midlands-based group as room booking was at a premium; ii) Small tests were run throughout the development process i.e. piloted in preliminery work, used for grounded theory generation, tested in the main research and modified in further format testing: iii) The cost per user was negligible at £2 and therefore came way under the '£25-£50' stated. Krug suggests what/how to test (N.B. 'Key task testing' denotes when users have been asked to do something/observations has been made as to how well they do):

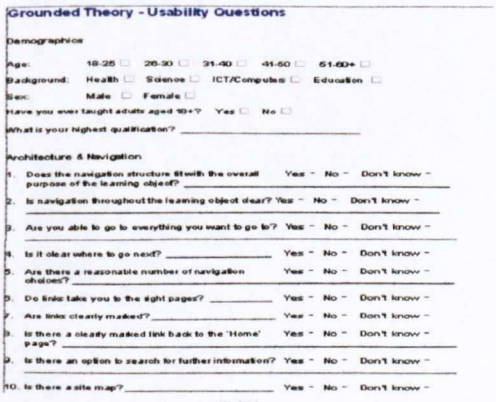
	Planning	Rough sketches	Page designs	Prototype	First usable version	Cubicle tests
What to test	Competitors sites	Home page sketch Top level site features/categories	Home page	Whatever is working	Whatever is working	Each page
Format	Live site	Paper	Paper	HTML	Live site	HTML page
How to test	Key tasks	Names of things	Basic navigation	Key tasks	Key tasks	Key tasks
What you're looking for	What do they like? What works well for their lives? Can they do key tasks?	Do they need/understand the site?	Can they find their way around the site?	Do they still 'get it'? Can they accomplish key tasks?	Do they still 'get it'? Can they accomplish key tasks?	Can they accomplish key tasks?
Session length	1 hr	15-20 min	15-20 min	45 min-1 hr	1 hr	5 min per page
No of tests	1	1-3	1-3	1-3	1-3	1 per page

APPENDIX 13 – Agimo's Usability Checklist from their User Profiling and Testing Toolkit

Available from: http://www.agimo.gov.au/publications/2004/06/toolkit/testing/checklist NB: This is not an exhaustive/complete test.

Design Component	Question
Architecture and Navigation	• Does the structure fit the purpose? • Is the navigation scheme clear? • Where are you? How do you find what you want? • Is it clear where to go next? • Does interaction support informational retrieval tasks? • Are there a reasonable number of nav-bar choices? • Do link names match page names? • Are links clearly marked? • Is there a clearly marked link back to the home page? • Is there an option to search for information? • Is there a site map? • Does every page have standard identification elements? • Does the user have control over navigation?
Layout and design	• Is layout consistent on all pages? • Is excessive scrolling required? • Is there a main display area on each page? • Does the layout work visually? • Proper use of alignment and grouping? • Proper use of contrast? • Is it cluttered?
Content	• Is the text clear and concise? • Is there excessive use of jargon or acronyms? • Is text organised in small chunks? • Are there spelling or grammar errors?
Forms	• Are dialog methods concise and consistent? • Are there clear submit and reset buttons?
Platform and implementation	• Is download time fast enough - do pages load in 3-10 seconds? • Do all the links work? • Are there broken images? • Are pages written to be found by search engines? • Does the site work with all supported browsers? • Does the site work on all platforms? • Does the homepage display fully on a screen with 800 x 600 resolution? • Does the site work on resolutions from 640 x 480 to 1024 x 768? • Are non-standard plug-ins used?
Accessibility (priority 1)	• Provide alternative text for all image type buttons in forms. • Provide text equivalents for every non-text element, including but not limited to images, scripts, animations, audio and video. • Ensure that all information conveyed with colour is also available without colour, for example from context or markup. • If an image conveys important information beyond what is in its alternative text, provide an extended description. • For tables not used as layout, such as spreadsheets, identify headers for the table rows and columns. • If the submit button is used as an image map, use separate buttons for each active region. • If a data table has two or more logical levels of row or column headers, use markup to associate data cells and header cells. • Provide alternative content for each SCRIPT that conveys important information or functionality

APPENDIX 14 – My Usability Checklist



Grounded Theory - Usability Questions

Demographics

Age: 18-25 ☐ 26-30 ☐ 31-40 ☐ 41-50 ☐ 51-60+ ☐

Background: Health ☐ Science ☐ ICT/Computers ☐ Education ☐

Sex: Male ☐ Female ☐

Have you ever taught adults aged 16+? Yes ☐ No ☐

What is your highest qualification? _____

Architecture & Navigation

1. Does the navigation structure fit with the overall purpose of the learning object? _____ Yes - No - Don't know -
2. Is navigation throughout the learning object clear? Yes - No - Don't know -
3. Are you able to go to everything you want to go to? Yes - No - Don't know -
4. Is it clear where to go next? _____ Yes - No - Don't know -
5. Are there a reasonable number of navigation choices? _____ Yes - No - Don't know -
6. Do links take you to the right pages? _____ Yes - No - Don't know -
7. Are links clearly marked? _____ Yes - No - Don't know -
8. Is there a clearly marked link back to the 'Home' page? _____ Yes - No - Don't know -
9. Is there an option to search for further information? Yes - No - Don't know -
10. Is there a site map? _____ Yes - No - Don't know -

11. Does every page have standard identification Elements (.i.e. look as if it belongs together?) Yes - No - Don't know -

12. Does the user have control over navigation? Yes - No - Don't know -

Layout and design

1. Do you like the overall 'look' of the learning object? Yes ☐ No ☐ Don't know ☐
2. What did you notice first on the initial page? _____
3. Is layout consistent on all pages? _____ Yes ☐ No ☐ Don't know ☐
4. Is excessive scrolling required? _____ Yes ☐ No ☐ Don't know ☐
5. Is there a main display area on each page? Yes ☐ No ☐ Don't know ☐
6. Is it cluttered anywhere in the learning object? Yes ☐ No ☐ Don't know ☐
7. What in your opinion makes a good learning package?
8. Does the learning object include these? Yes ☐ No ☐ Don't know ☐
   To what extent? _____
9. What elements (i.e. text, narration, etc) in the learning object do you like best?
10. Do you think a combination of elements operating together (i.e. text, narration, animation, images) are better than single elements? Yes ☐ No ☐ Don't know ☐
    Please state reasons _____

Content

1. Is the text clear and concise? _____ Yes ☐ No ☐ Don't know ☐
2. Is there excessive use of jargon or acronyms? ____ Yes ☐ No ☐ Don't know ☐
3. Is text organised in small chunks? _____ Yes ☐ No ☐ Don't know ☐
4. Are there spelling or grammar errors? _____ Yes ☐ No ☐ Don't know ☐

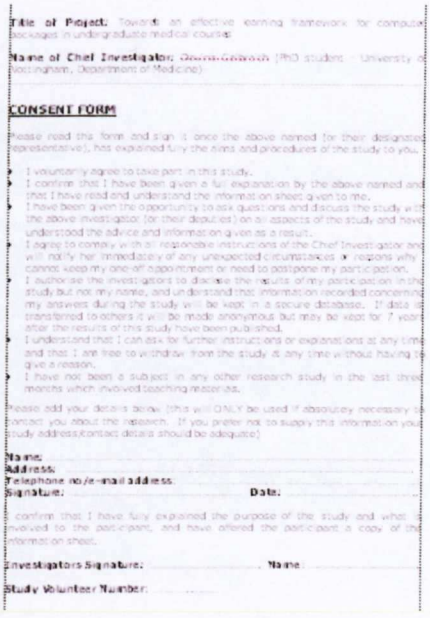
Platform and implementation

1. Does the learning object load properly? _____ Yes ☐ No ☐ Don't know ☐
2. Do images load within 3-10 seconds? _____ Yes ☐ No ☐ Don't know ☐
3. Does the WAP connect properly? _____ Yes ☐ No ☐ Don't know ☐
4. Does the initial page display fully? _____ Yes ☐ No ☐ Don't know ☐
5. Are there broken images? _____ Yes ☐ No ☐ Don't know ☐
6. Do all the links work? _____ Yes ☐ No ☐ Don't know ☐
7. Are pages found easily? _____ Yes ☐ No ☐ Don't know ☐

Please put any other comments you may have here:

APPENDIX 15 – Participant Information Sheet and Consent Form – Usability Testing

The consent form for the Midlands site is shown below:



Title of Project: Towards an effective learning framework for computer packages in undergraduate medical courses

Name of Chief Investigator: [redacted] (PhD student – University of Nottingham, Department of Medicine)

CONSENT FORM

Please read this form and sign it once the above named (or their designated representative), has explained fully the aims and procedures of the study to you.

- I voluntarily agree to take part in this study.
- I confirm that I have been given a full explanation by the above named and that I have read and understand the informat on sheet given to me.
- I have been given the opportunity to ask questions and discuss the study with the above investigator (or their deputies) on all aspects of the study and have understood the advice and informat on given as a result.
- I agree to comply with all reasonable instructions of the Chief Investigator and will notify her immediately of any unexpected circumstances or reasons why I cannot keep my one-off appointment or need to postpone my participation.
- I authorise the investigators to disclose the results of my participation in the study but not my name, and understand that information recorded concerning my answers during the study will be kept in a secure database. If data is transferred to others it will be made anonymous but may be kept for 7 years after the results of this study have been published.
- I understand that I can ask for further instructions or explanations at any time and that I am free to withdraw from the study at any time without having to give a reason.
- I have not been a subject in any other research study in the last three months which involved teaching research.

Please add your details below. (this will ONLY be used if absolutely necessary to contact you about the research. If you prefer not to supply this information your study address/contact details should be adequate)

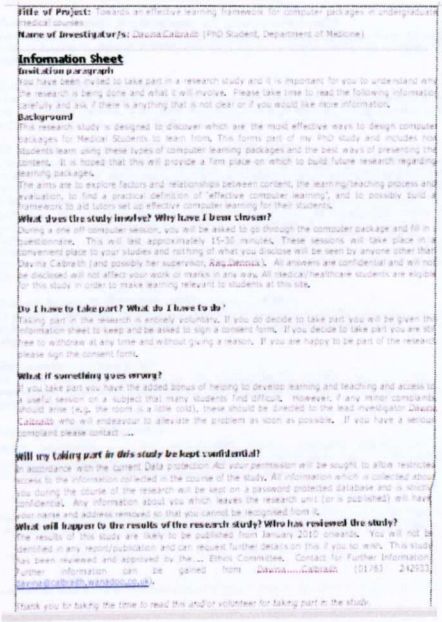
Name:
Address:
Telephone no/e-mail address:
Signature: Date:

I confirm that I have fully explained the purpose of the study and what is involved to the participant, and have offered the participant a copy of the informat on sheet.

Investigators Signature: Name

Study Volunteer Number:

The Participant Information Sheet for the Midlands site is shown below:



Title of Project: Towards an effective learning framework for computer packages in undergraduate medical courses

Name of Investigator/s: Davina Calbraith (PhD Student, Department of Medicine)

Information Sheet

Invitation paragraph

You have been invited to take part in a research study and it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and ask if there is anything that is not clear or if you would like more information.

Background

This research study is designed to discover which are the most effective ways to design computer packages for Medical Students to learn from. This forms part of my PhD study and includes how students learn using these types of computer learning packages and the best ways of presenting the content. It is hoped that this will provide a firm place on which to build future research regarding learning packages.

The aims are to explore factors and relationships between content, the learning/teaching process and evaluation, to find a practical definition of 'effective computer learning', and to possibly build a framework to aid tutors set up effective computer learning for their students.

What does the study involve? Why have I been chosen?

During a one off computer session, you will be asked to go through the computer package and fill in a questionnaire. This will last approximately 15-30 minutes. These sessions will take place in a convenient place to your studies and nothing of what you disclose will be seen by anyone other than Davina Calbraith (and possibly her supervisor, Reg Dennick). All answers are confidential and will not be disclosed nor will not affect your work or marks in any way. All medical/healthcare students are eligible for this study in order to make learning relevant to students at this site.

Do I have to take part? What do I have to do?

Taking part in the research is entirely voluntary. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. If you are happy to be part of the research please sign the consent form.

What if something goes wrong?

If you take part you have the added bonus of helping to develop learning and teaching and access to a useful session on a subject that many students find difficult. However, if any minor complaints should arise (e.g. the room is a little cold), these should be directed to the lead investigator Davina Calbraith who will endeavour to alleviate the problem as soon as possible. If you have a serious complaint please contact ....

Will my taking part in this study be kept confidential?

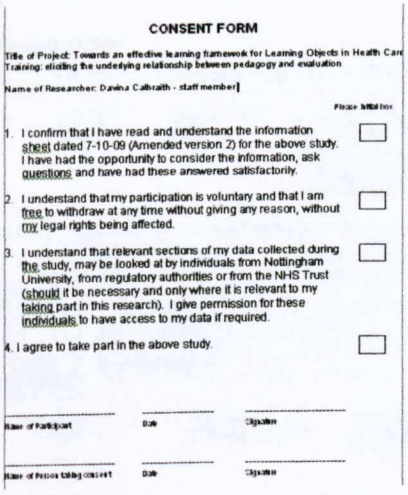
In accordance with the current Data protection Act your permission will be sought to allow restricted access to the information collected in the course of the study. All information which is collected about you during the course of the research will be kept on a password protected database and is strictly confidential. Any information about you which leaves the research unit (or is published) will have your name and address removed so that you cannot be recognised from it.

What will happen to the results of the research study? Who has reviewed the study?

The results of this study are likely to be published from January 2010 onwards. You will not be identified in any report/publication and can request further details on this if you so wish. This study has been reviewed and approved by the ... Ethics Committee. Contact for Further Information: Further information can be gained from Davina Calbraith (01783 [illegible]) (davina@calbraith.wanadoo.co.uk).

Thank you for taking the time to read this and/or volunteer for taking part in the study.

Consent form for the Eastern site:



CONSENT FORM

Title of Project: Towards an effective learning framework for Learning Objects in Health Care Training: eliciting the underlying relationship between pedagogy and evaluation

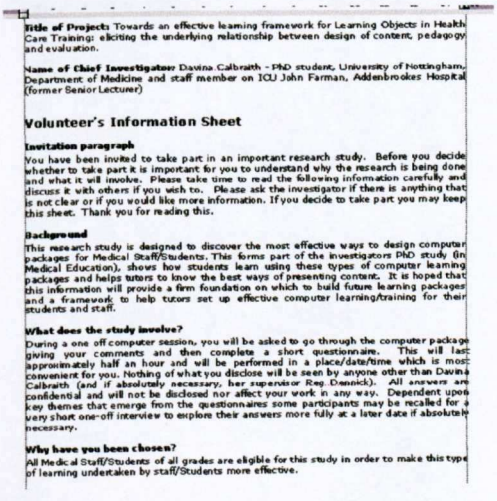
Name of Researcher: Davina Calbraith - staff member

Please initial box

1. I confirm that I have read and understand the information sheet dated 7-10-09 (Amended version 2) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my legal rights being affected. ☐
3. I understand that relevant sections of my data collected during the study, may be looked at by individuals from Nottingham University, from regulatory authorities or from the NHS Trust (should it be necessary and only where it is relevant to my taking part in this research). I give permission for these individuals to have access to my data if required. ☐
4. I agree to take part in the above study. ☐

Name of Participant	Date	Signature
Name of Person taking consent	Date	Signature

Information Sheet for the Eastern site:



Title of Project: Towards an effective learning framework for Learning Objects in Health Care Training: eliciting the underlying relationship between design of content, pedagogy and evaluation.

Name of Chief Investigator: Davina Calbraith - PhD student, University of Nottingham, Department of Medicine and staff member on ICU John Farman, Addenbrookes Hospital (former Senior Lecturer)

Volunteer's Information Sheet

Invitation paragraph

You have been invited to take part in an important research study. Before you decide whether to take part it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish to. Please ask the investigator if there is anything that is not clear or if you would like more information. If you decide to take part you may keep this sheet. Thank you for reading this.

Background

This research study is designed to discover the most effective ways to design computer packages for Medical Staff/Students. This forms part of the investigators PhD study (in Medical Education), shows how students learn using these types of computer learning packages and helps tutors to know the best ways of presenting content. It is hoped that this information will provide a firm foundation on which to build future learning packages and a framework to help tutors set up effective computer learning/training for their students and staff.

What does the study involve?

During a one off computer session, you will be asked to go through the computer package giving your comments and then complete a short questionnaire. This will last approximately half an hour and will be performed in a place/date/time which is most convenient for you. Nothing of what you disclose will be seen by anyone other than Davina Calbraith (and if absolutely necessary, her supervisor Reg Dennick). All answers are confidential and will not be disclosed nor affect your work in any way. Dependent upon key themes that emerge from the questionnaires some participants may be recalled for a very short one-off interview to explore their answers more fully at a later date if absolutely necessary.

Why have you been chosen?

All Medical Staff/Students of all grades are eligible for this study in order to make this type of learning undertaken by staff/Students more effective.

Do I have to take part? What do I have to do?
Taking part in the research is entirely voluntary. You can decide whether or not to take part. If you decide to take part you may keep this information sheet and you will be asked to sign a consent form and make a date/time to go through the package. (N.B. You are free to withdraw at any time and without giving a reason).

What are the possible disadvantages and risks of taking part?
There are no possible adverse effects of taking part! If you take part you have the added bonus of helping to develop learning and teaching and a useful session on a subject that many students find difficult.

What if something goes wrong?
If any complaints should arise, these will be handled initially by the Chief Investigator and if no satisfactory outcome is achieved then they should be directed to Reg Dennick at the Medical Education Unit, B Floor, Medical School, Queen's Medical Centre, Nottingham).

Will my taking part in this study be kept confidential?
In accordance with the current Data protection Act all information which is collected from you during the course of the research will be anonymised using numbers only and will be kept on a confidential password protected database in a locked cupboard. It will be used only for the purposes of the research only in order to develop further computer packages. Your permission will be sought to allow restricted access (i.e. the researcher and/or the researcher's supervisor ONLY) to the information collected in the course of the study. Any information about you which leaves the research unit (or is published) will have all information about you removed so that you cannot be recognised from it. Data will be kept for 7 years (in accordance with research ethical and data protection requirements), *after which time it will be shredded and disposed of securely.*

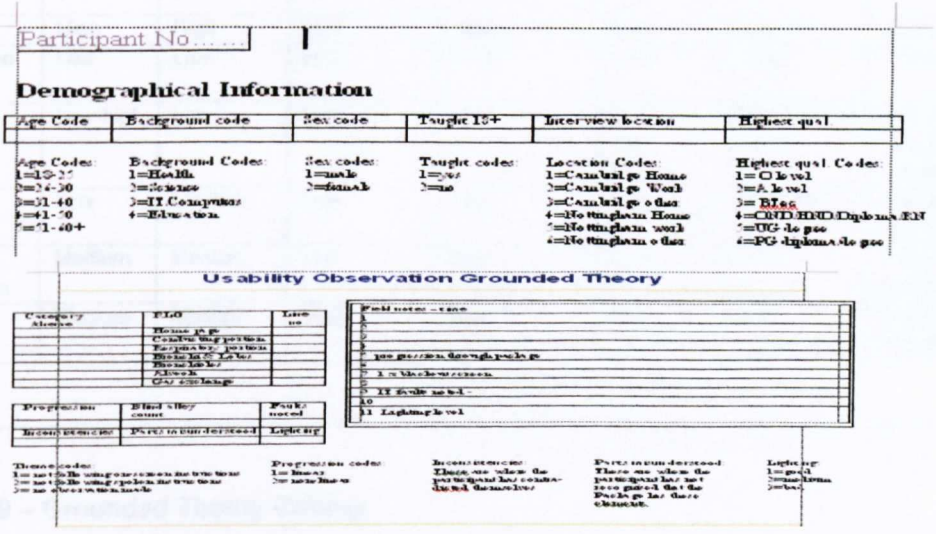
What will happen to the results of the research study?
The results of this study are likely to be published from the end of 2009 onwards. You will not be identified in any report/publication and can request further details on this if you so wish.

Who has reviewed the study?
This study has been approved by the Cambridgeshire 3 REC, the Research and Development Department, a Peer Review Supervisor, and also the ICU Ward Manager/Lead Clinical Nurse.

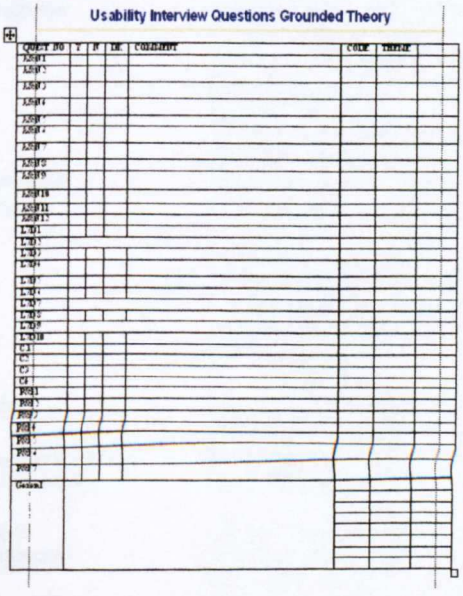
Contact for Further Information
Further information can be gained from Davina Calbraith (01763 242933, davina@calbraith.wanadoo.co.uk), Medical Education Unit, B Floor, QMC. Or alternatively from Dr Reg Dennick Medical Education Unit, B Floor, QMC.

Thank you for taking the time to read this and/or volunteer for taking part in the study. You will be given a copy of the information sheet and a signed consent form to keep.

APPENDIX 16 - Usability Observation Fieldwork Sheet Excerpts



APPENDIX 17 - Data Collation Sheet



APPENDIX 18 - Grounded Theory Stages and Coding

The stages and coding were used as follows:

Stage 1 - Getting to know the raw data
Stage 2 - Giving conceptual labels to the phenomena, grouping those concepts into broader categories/themes **(Open coding)**
Stage 3 - Reconstructing data by exploring linkages between categories. Analysed in terms of causality, context and interactions **(Axial Coding)**
Stage 4 - Create core categories to which other categories are systematically related **(Selective Coding)**
Stage 5 - Relate the core category to other phenomena and contexts **(Conditional Matrix)**

An example of this is given below (this is based on Ryan & Bernard, 2003, Field Methods, Sage publications, Vol.15, No.1, pp85–109) and can be seen in Chapter 3, 'Finding themes':

Technique	Labor Intensity	Language Expertise	Substantive Expertise	Methodology Expertise	Stage of Analysis	Number of Themes produced	Type of theme produced
Repetitions	Low	Low	Low	Low	Early	High	Theme
Indigenous typologies	Low	High	Low	Low	Early	Medium	Theme, subtheme
Metaphors	Low	High	Low	Low	Early	Medium	Theme
Transitions	Low	Low	Low	Low	Early	High	Theme
Similarities and differences	Low-high	Low	Low	Low	Early	High	Theme
Linguistic connectors	Low	High	Low	Low	Late	High	Theme
Missing data	High	High	High	High	Late	Low	Theme
Theory-related material	Low	Low	High	High	Late	Low	Theme
Cutting and sorting	Low-high	Low	Low	Low	Early or late	Medium	Theme, subtheme, metatheme
Word lists and KWIC	Low	Medium	Low	Low	Early	Medium	Theme, subtheme
Word co-occurrence	Medium	Medium	Low	High	Late	Low	Theme, metatheme
Metacoding	Medium	Medium	High	High	Late	Low	Theme, metatheme

APPENDIX 19 – Grounded Theory Coding

Coding for the usability study was as follows:

1= participant not following on-screen instructions
2= participant not following verbal instructions
3= no observation made
4= good layout
5= too much information
6= participant noticed colours first
7= good labels
8= good images/pictures
9= participant prefers summarised information
10= participant prefers simple terminology

21= navigation needs an overview
22= good text
23= participant noticed image first
24= bad layout
25= learning channel
26= good audio
27= changes wanted: more funny images
28= participant noticed text first
29= participant prefers 2+ elements to ↑ understanding
30= participant prefers 2+ elements to ↓ boredom

41= participant prefers animation elements
42= participant finds mixed elements engaging
43= the colours are good for dyslexia
44= changes wanted: definition

11= participant prefers simple elements
12= participant prefers colourful elements
13= participant prefers 1 element per screen
14= participant prefers interaction
15= participant prefers 2 elements per screen
16= bad font
17= font is bad contrast with background
18= good speed of text revelation
19= pt learns using 'revisit material' options
20= navigation is unclear/confusing

31= learning is not foolproof!
32= changes wanted: content list
33= changes wanted: page numbers
34= changes wanted: normal x-ray
35= changes wanted: more labels
36= changes wanted: more learning objectives
37= participant prefers bullet point elements
38= participant prefers image/picture elements
39= participant prefers test elements
40= participant prefers narration elements

51= no choice regarding navigation
52= participant prefers problem-solving elements
53= participant prefers easy-to-follow elements
54= participant prefers test opportunity elements

45= changes wanted: more animations
46= changes wanted: more images/pictures
47= changes wanted: VQ information slower
48= changes wanted: better animation
49= changes wanted: better layout
50= 'clumsy' navigation

61= good choice of navigation
62= participant prefers 'humurous' elements
63= participant noticed symmetrical layout first
64= bad text - too much
65= learning objectives come up too fast
66= changes wanted: more information
67= changes wanted: more audio
68= participant prefers clear elements
69= changes wanted: more internet links
70= participant prefers all elements

81= good learning objects
82= good information
83= participant unable to link information
84= good explanation
85= inconsistent text font
86= 'blocks' of text
87= slide is 'unfocused' (fuzzy)
88= image blocks text
89= good animation
90= participant missed these entirely

55= participant prefers pop-up boxes for glossary
56= changes wanted: 'click to enlarge facility'
57= information gives knowledge
58= participant prefers mixed elements
59= layout needs direstion arrow
60= good navigation

71= learning styles
72= changes wanted: more highlighted text
73= participant prefers 'click to move'elements
74= changes wanted: more detailed/better pictures
75= changes wanted: simple/clear/easy to follow elements
76= audio not 'synch'd'
77= participant had different expectation
78= participant had expectation fulfilled
79= information is well-timed
80= layout provides 'good' stimuli

91= participant found punctuation irritating
92= bad/inconsistent layout
93= image is bad/unclear/fuzzy
94= learning is inconsistent
95= participant noticed image and layout first

APPENDIX 20 - Categorisation Of Codes

Usability study findings were coded as in Appendix 19 and was then grouped according to subject and whether the statement was positive, neutral or negative:

Information

Positive comments	Neutral comments	Negative comments
57= information gives knowledge		5 = too much information
79= information is well-timed		83= pt unable to link information
82= good information		
84= good explanation		

Layout

Positive comments	Neutral comments	Negative comments
4 = good layout		24= bad layout
		59= layout needs direction arrow
		92= inconsistent layout

Audio

Positive comments	Neutral comments	Negative comments
26= good audio		76= bad audio/not 'synch'd'

Text

Positive comments	Neutral comments	Negative comments
7 = good labels		16= bad font/hard to read
18= good speed of text revelation		17= font is bad contrast with background
22= good text		64= bad text - too much
		85= inconsistent text font
		86= 'blocks' of text

Images

Positive comments	Neutral comments	Negative comments
8 = good images/pictures		87= slide is 'unfocused' (fuzzy)
88= image blocks text		
93= image is bad/confused		

Animation

Positive comments	Neutral comments	Negative comments
89= good animation		

Observations

Positive comments	Neutral comments	Negative comments
	3= no observation made	1= participant not following on-screen instructions
2= participant not following verbal instructions		90= participant missed these entirely
		91= participant found punctuation irritating

Learning/pedagogy

Positive comments
25= learning channel
43= colours good for dyslexia
71= learning styles
81= good learning objects
78= pt had expectation fulfilled

Neutral comments
77= pt had different expectation

Negative comments
19= pt only learns using 'revisit material' options
31= learning is not foolproof!
65= learning objectives come up too fast
94= learning is inconsistent

First thing noticed

Positive comments

Neutral comments
6 = participant noticed colours first
23= participant noticed image first
28= participant noticed text first
63= participant noticed symmetrical layout first
95= participant noticed image layout first

Negative comments

Navigation

Positive comments
60= good navigation
61= good choice of navigation

Neutral comments

Negative comments
20= navigation is unclear/confusing
21= navigation needs an overview
50= 'clumsy' navigation
51= no choice regarding navigation

Preferred elements

Positive comments

Neutral comments

Element:
11= participant prefers simple elements
12= participant prefers colourful elements
13= participant prefers 1 element per screen
15= participant prefers 2 elements per screen
29= participant prefers 2+ elements to ↑ understanding
30= participant prefers 2+ elements to ↓ boredom
42= participant finds mixed elements engaging
58= participant prefers mixed elements
68= participant prefers clear elements

Written:
9 = participant prefers summarised information
10= participant prefers simple terminology
37= participant prefers bullet point elements

Visual:
38= participant prefers image/picture elements
41= participant prefers animation elements

Audio:
40= participant prefers narration elements

Other:
14= participant prefers interaction
39= participant prefers test elements
52= participant prefers problem-solving elements
53= participant prefers easy-to-follow elements
54= participant prefers test opportunity elements
55= participant prefers pop-up boxes for glossary
62= participant prefers 'humurous' elements
63= participant noticed symmetrical layout first
73= participant prefers 'click to move'elements

Negative comments

Changes wanted

Positive comments

Neutral comments

Visual:
27= changes wanted: more funny images
34= changes wanted: normal x-ray
45= changes wanted: more animations
46= changes wanted: more images/pictures
48= changes wanted: better animation
49= changes wanted: better layout
56= changes wanted: 'click to enlarge facility'
69= changes wanted: more internet links
74= changes wanted: more detailed/better pictures

Written:
32= changes wanted: content list
33= changes wanted: page numbers
35= changes wanted: more labels
36= changes wanted: more learning objectives
44= changes wanted: definition

Negative comments

47= changes wanted: VQ information slower
66= changes wanted: more information
72= changes wanted: more highlighted text
75= changes wanted: simple/clear/easy to follow elements

Audio:
67= changes wanted: more audio

APPENDIX 21 — Generic Systematic Review Protocol

As the protocol is too large to be included here please see separate bound protocol (available on request from the author).

APPENDIX 22 - Specific Data Collection Questions for LO Research Systematic Review

No unique questions were needed for the LO research systematic review.

APPENDIX 23 - Specific Data Collection Questions for LO Pedagogy Systematic Review

Specific features of the pedagogy e/g/ its constituent parts, advantages, disadvantages, whether it simulates clinical practice/physiological processes etc. (This is available on request from the author).

APPENDIX 24 - Specific Data Collection Questions for LO Evaluation Systematic Review

Type of evaluation used and closely it keeps to its ascribed approach, specific features of evaluations e.g. its constituent parts, advantages, disadvantages, whether it simulates clinical practice/physiological processes etc. (This is available on request from the author).

APPENDIX 25 – Generic Data Extraction Sheets

(These are available on request from the author).

APPENDIX 26 - Methodological Rigor Sheets

To ensure that the systematic reviews gather empirical data of high rigor (both quantitative and the more non-comparative.or qualitative type studies) 'methodological rigor sheets' were devised by the Researcher (based on BEME criteria). (These are available on request from the author).

APPENDIX 27 - Kirkpatrick Measurement of Impact

Kirkpatrick hierarchy:

Level 1 – Participation (Learners views on the learning experience, its organisation, presentation, content, teaching methods, and aspects of the instructional organisation, materials, quality of instruction)
Level 2a - Modification of Attitudes/perceptions (Changes in the reciprocal attitudes/perceptions between participant groups toward intervention/simulation)
Level 2b - Modification of knowledge/skills (Knowledge – acquisition of concepts, procedures and principles. Skills – acquisition of thinking/problem-solving, psychomotor and social skills)
Level 3 - Behavioural change (Documented transfer of learning to the workplace or willingness of learners to apply new knowledge and skills)
Level 4a - Change in organisational practice (Wider changes in the organisational/delivery of care, attributable to an educational program)
Level 4b - Benefits to patients/clients (Any improvement in health of patients as a direct result of an educational programme)

APPENDIX 28 - Systematic Review Evaluation Sheet

I devised this in 2005 with guidelines based on EPPI review 2002. (This is available on request from the author).

APPENDIX 29 – SCIE's Systematic Review Evaluation Sheet 1

SCIE published systematic review evaluation guidelines. This evaluation sheet is from the SCIE Knowledge review booklet guidelines - 'The Conduct of Systematic Research Reviews for SCIE Knowledge reviews'.

Systematic Review Evaluation Sheet 1

Taken from SCIE Knowledge review booklet guidelines - 'The Conduct of Systematic Research Reviews for SCIE Knowledge reviews'.

A) PRELIMINARIES - To be completed for all types of study:

1. P17 - Section 92 - Inclusion/exclusion criteria for effectiveness/ intervention should be based on the folloing areas therefore ask these questions:

- Participant/user group - who?
- Intervention type - what?
- Intervention setting - where?
- Intervention provider - who?
- Outcomes to be considered - what?
- Types of studies to be reviewed - what?

2. P29 - Section 147 - In order to pick up **limitations of quality appraisal** in this type of study ask the following questions:

	Yes	No	D/K	N/A
Sample/ling				
• Appropriate/justified selection of cases/participants?	□	□	□	□
• Adequate sample description?	□	□	□	□
Data collection				
• Right questions asked?	□	□	□	□
• Good follow up?	□	□	□	□
Data analysis				
• Searched for negative cases?	□	□	□	□
• Good use of data? (i.e. not selective)	□	□	□	□
Data analysis products				
• Variation in theory/explanatory concepts?	□	□	□	□
• Interpretations fit data?	□	□	□	□

3. P31 - Section 157 - **Assessing levels of relevance:**

	Yes	No	D/K	N/A
• Is the overall focus of the study relevant to the review?	□	□	□	□
• Is the concpetual focus of the study relevant to the review?	□	□	□	□
• Is the theoretical focus of the study relevant to the review?	□	□	□	□
• Is the context of the study relevant to the review?	□	□	□	□
• Is the sample of the study relevant to the review?	□	□	□	□
• Are the outcomes measured relevant to the review?	□	□	□	□
• Are the methods of measurement relevant to the review?	□	□	□	□

3. P32 - Section 161 - **Quality markers**

	Yes	No	D/K	N/A
• Strength of design - did authors report material Relevant to research?	□	□	□	□
• Did authors support general perpectives already known? (my interpretation of what they are saying here!)	□	□	□	□
• Is enough depth given to assess quality of reporting? (or analysis? Thereby giving confidence?)	□	□	□	□
• Has study assessed whether it is generalisable? (to the wider population?)	□	□	□	□

B) SYNTHESIS

For Quantitative OR Empirical Qualitative studies **ONLY:**

P37 - Section 185 - Narrative synthesis (quant & qual- but EMPIRICAL ONLY). Based on Popay et al 2006 'Guidance on the conduct of Narrative Synthesis, Lancaster Univeristy. The following must be carried out AND NOTED HERE for synthesis:

- A description (of included study)
- A description of findings
- A description of synthesis
- A 4 step sequnetial process (i.e. Narrative sysnthesis framework as outlined in Pay J, Robert H, Sowden A, Petigrew M, AraiL, Roen K, Rogers M 2004 'Developing guidance on the conduct of narrative synthesis' Journal of Epidemiology and

Community Health Vol. 59 Supplement 1:A7 - cited in p38 'The Conduct of Systematic Research Reviews for SCIE Knowledge reviews') including develop a theory of how, why and for whom the intervention works, thereby aiding judgements regarding research question, inclusion criteria and study findings;

1. preliminary synthesis of included studies findings in order to describe patterns accross studies;
2. exploring relationships in the data in order to find factors that may explain differences accross studyies findings;
3. assessing robustness of the synthesis in order to judge the strength of evidence in the review.

For non-empirical qualitive data **ONLY:**

P38/39 - Section 188, 189 & 190 - QUALITATIVE DATA SYNTHESIS - i.e. to develop conceptual synthesis. **This has 3 stages:**

i) 'First-order finding/interpretations' - i.e. the meanings reported to researchers;
ii) Second-order interpretations' - i.e. the constructs/interpretations that researchers place on i)
iii) 'Third-order interpretations' - i.e. explanations and hypotheses developed by reviewers arising out of ii).

This should be done by means of the following and noted down here:

- identify key concepts (can use Nvivo, RevMan or Nudist etc to help here to gain conceptual categories and the studies that support them - similar to grounded theory;
- compare core findings and concepts accross studies (this process is sometimes called 'translation' or 'reciprical translational anaysis', grounded theory approach is relevant as the process resembles that of seeking similarities/differences between findings and concepts and why this is so - e.g. 'deviant case analysis' - also akin to 'refutational analysis in QDS) in order to test general conceptsuntil it is clear that they are central ones;
- develop line of argument/reasoned case (to link findings and concepts) and tabulated (as shown on pages 44-46) showing the first, second and third order stages.

For mixed methods data **ONLY:**

P39 - Section 191 - developed by the EPPI centre - i) systematic review synthesis, ii) qual synthesis of service users regarding the intervention, iii) synthesis of i) and ii) thereby fusing all together to identify directions for practice development (see SCIE Conduct of systematic research reviews for SCIE knowledge reviews publication - i.e. a PRACTICE SURVEY CAN BE UNDERTAKEN See p41 for details - and the two parts fused together)

C) DISCUSSION - FOR ALL STUDIES:

P43 - Section 207 - (Higgins JPT, Green S - Eds, Cochrane Handbook for Systematic Reviews of interventions, 17-05-2000, Cochrane library, issue 4.2.5, Chichester, Wiley) - 4 dimensions should be considered when discussing results:

- strength of evidence
- applicability fo the results
- consideration of current practice/cost etc
- important trade-offs between expected benefits/harms of the intervention

D) ANALYSIS/EVALUATION OF REVIEW PROCESSES - FOR ALL STUDIES

P45 - Section 217, 218 & 219

Impact of carer/user involvement:

User Participation
Methods of involvement
Specific user contributions
User views
User feedback

Other review processes reflections on
Quality appraisal
Synthesis
Other miscellaneous processes

E) LIMITATIONS - FOR ALL STUDIES:

P43 - Section 213 - This includes limitations to the:

- review teams 'representativeness'
- search
- overall study due to gaps in evidence base

F) IMPLICATIONS - FOR ALL STUDIES:

P44 - Section 214 – For users? Carers? Practice? Policy? Research?

G) SUMMARY OF FINDINGS - FOR ALL STUDIES:

P27 - Section 138 summary table (When writing this up set out in table from as in Pettigrew M and Roberts H 2006 Systematic reviews in the social sciences; a practical guide' Oxford, Blackwell).

Study characteristics:

Sample - number of participants
- ethnicity "
- age "
- gender "

job/discipline "
Intervention - who delivered it?
Frequency
Intensity
Outcomes measured
Research method
Findings summary

Now do Stage II Phase 7 from CRD document 'Undertaking systematic review's of research on effectiveness' document for in depth further analysis - i.e. *Systematic Review Evaluation Sheet 2*

APPENDIX 30 – SCIE's Systematic Review Evaluation Sheet 2

Evaluation sheet 2 is taken from three sources: EPPI REVIEW, June 2002, EPPI-Centre - 'A systematic review of the impact of summative assessment and tests on students' effectiveness for learning' and Stage II Phase 7 (chapter 6 & 7) CRD document. NB the word 'motivation' appears (and has been replaced here) in the following text by me with the word 'effectiveness':

IN-DEPTH REVIEW: RESULTS

6.1 Description of included studies

Table 4: Weight of evidence for each study taken from Systematic Review Evaluation Sheet 1:- 0-33%=Low; 34-66%=Medium; 67-100%=High

Study No.	Type of study	A Basics complete	B Appropriateness of sample design, data collection/analysis	C Relevance of topic focus to review	D Quality	E Overall weight of evidence (check whether A-D are SCIE-based)

There were studies of high, medium and low weight and those yet to be classified.

Table 5: Type of approach for the included studies

Reference	Pedagogical approach	Paradigm	ISD approach	Hardware	Software	Learning theory approach

Table 6: Findings, concepts & hypotheses (first 5 columns + no of categories in column 6)

Ref	First order findings	Second order findings	Third order findings	Main concept	Concept & findings	Reasoned argument/ hypothesis	Level of robust argument

There were studies of high, medium and low weight and those that had insufficient details.

Issue	Important considerations
Choice of effect measure	• The type of data to be analyzed (binary, continuous, time-to-event) • The consistency of estimates of the effect measure across studies • The ease of interpretation of the chosen effect measure
Choice of statistical model	• The reliability of the method • when sample sizes are small • when events are rare • The degree of between-study heterogeneity
Heterogeneity of effect measures between studies	• The assessment of the degree of heterogeneity • If heterogeneity is substantial • whether formal meta-analysis should be considered, • whether an overall summary has a sensible meaning, • whether random-effect or other multilevel modelling approaches should be used to account for between-study heterogeneity • whether the impact of other study level factors on effect measure can be investigated in stratified analyses or meta-regression models

---

APPENDIX 32 – Types of Learning Object Research Approaches Needed for Main Study (SRES2 - RIGOR)

HIGH
Study 3017 - non-comparative
Study 1016 - non-comparative
Study 1018 - review

MEDIUM/HIGH
Study 1013 - non-comparative
Study 1004 - non-comparative
Study 1003 - review

MEDIUM/LOW
Study 1030 - non-comparative
Study 3001 - non-comparative
Study 1006 - review

LOW
Study 1017 - non-comparative
Study 1008 - comparative CBA

APPENDIX 33 – Types of Approaches Needed for Main Study (SRES2 - RIGOR)

LO Evaluation

HIGH
Utilisation-focused - Patton 1887 & 2001 (study 8007)
Own-type used (study 8002)
Not stated (Study 8004)
Not stated (Study 8001)

MEDIUM/HIGH
Other - Finne et al 1995 Trailing Evaluation Methodology (study 8013)
Utilisation-focused - Patton 1987 & 2001 + CIPP - Shufflebeam 1971 (calls it participant-oriented - study 8019)
Convergent Participant – Nesbit et al (study 8008)
Democratic - House & Howe 1999 (study 8019)
Democratic - Ryan & Stephens 2000 (study 8019)
Not stated (Study 8010)

MEDIUM/LOW
Collaborative - Cousins et al 2002 Communications Model (study 8015)
Collaborative - Cecezkexmanovic & Webb 2000 Communicative Model of Collaborative (study 8015)
Collaborative - Learning & Habermas 1984 theory of communicative action (study 8015)
Not stated (study 8005)

Not stated (study 8006a)
Not stated (study 8009)

LOW
Layered - Kara, Giannidis et al 2001 (study 8006b)
Not stated (Study 8012)
Not stated (Study 8014)
Not stated (Study 8017)

Unknown
N/A as there were none in this category

Pedagogy

HIGH
Concept Dev Theory - Dickelman et al 1999
Concept Dev Theory - Dickelman et al 1999 + Active Learning Theory - Hall Hutchings & White/Halliday 1973&1975
Concept Dev Theory - Dickelman et al 1999 + Active Learning Theory - Gagne 1985 & 1992
Concept Dev Theory - Dickelman et al 1999 + Emancipatory
Good UG Principles - Chickering & Gamson 1989 + Cognitive developmental - Piaget 1970
Good UG Principles - Chickering & Gamson 1989 + Active Learning Theory - Hall Hutchings & White/Halliday 1973 & 1975
Good UG Principles - Chickering & Gamson 1989 + Active Learning Theory - Gagne 1985 & 1992

MEDIUM/HIGH
Active Learning Theory - Hall Hutchings & White/Halliday 1973&1975 (
Active Learning Theory - Boyle & Cook 2002/Halliday 1973 & 1975
Cognitive developmental - Piaget 1970
Situated learning - Wiley et al 2003
Situated learning - Laurillard 2002

MEDIUM/LOW
SOI/Dimensional Hyperlearning model - Billings 2000
SOI/Dimensional Hyperlearning model - Jeffries 2000
Situated learning - Hodges & Sasnett 1993

LOW
Active Learning Theory - Gagne 1985 & 1992
Learning Object Design and Sequencing Theory (LODAS) - Wiley 2000 & 2001
Learning Object Design and Sequencing Theory (LODAS) - T Reeves YR
Experiential Learning Theory - Bruner 1975
Experiential Learning Theory - Santrook 2001
Behavioural Connectivist Learning Theory - Pavlov YR
Behavioural Connectivist Learning Theory - Thorndike, YR
Behavioural Connectivist Learning Theory - Skinner, YR
Behavioural Connectivist Learning Theory - Rappan 1998
Behavioural Connectivist Learning Theory - Watson YR
Emancipatory Model Theory - Mayes & Fowler 1999
Emancipatory Model Theory - Taylor 1980
Emancipatory Model Theory - Rogers YR
Emancipatory Model Theory - Vygotsky 1962 situated learning
Emancipatory Model Theory - Reese & Overton Elemental model 1970 + Knowles et al 1998 Holistic model
Emancipatory Model Theory - Rusby 1979

Unknown/unrecordable
SOI/Dimensional Hyperlearning model - Meyer 1996
SOI/Dimensional Hyperlearning model - Shutzman YR
Constructivist Learning Environments (CLE) - Jonassen 1999
Situated learning - Herrington & Oliver 2000
Emancipatory Model Theory - Vygotsky 1962 zone of prox. Dev.
Component Display Theory (CDT) - Merrill YR

APPENDIX 34 – Types of Learning Object Research Approaches Needed for Main Study (SRES2 — *Research Design Components*)

HIGH
Study 3017 - observation (evidence with data)
Study 1016 - single group (time interrupted series)
Study 1018 - review (conceptual)

MEDIUM/HIGH
Study 1013 - case-series & survey (conceptual)
Study 1004 - observation (conceptual)
Study 1003 - review (analysis & evidence with data)

MEDIUM/LOW
Study 1030 - case series (evidence with data)
Study 3001 - observation & case study (evidence with data)
Study 1006 - review (?)

LOW
Study 1017 - observation (evidence with limited data)
Study 1008 - single group (before & after contemporaneous CBA)

APPENDIX 35 – Data Sources

HIGH
Study 3017 - student survey, groupwork poster presentation, student self assessment, practice exam with feedback, weekly MCQ quiz, lab experiments
Study 1016 - MCQ exam
Study 1018 - opinion literature and experience

MEDIUM/HIGH
Study 1013 - student survey, interview (face to face & telephone)
Study 1004 - course evaluation, observation, e-mail/discussion postings analysis
Study 1003 - survey

MEDIUM/LOW
Study 1030 - questionnaire (LORI - students & faculty), student survey, verbal (students), student usability testing.
Study 3001 - questionnaire
Study 1006 - student survey, case study, interview, focus group, vignettes, HEI documents, literature on VETBAC project (Whithear et al 1994/McNaught et al 2002)

LOW
Study 1017 - questionnaire, interview, tracking devices, observation, no of passes
Study 1008 - questionnaire, interview, tracking device

Here the 2 studies with MCQs score high, and the 2 studies with 'questionnaire, interview, tracking device' mix score low. Other than this no real patterns can be seen.

APPENDIX 36 - Stated Expected Outcomes

HIGH
Study 3017 - increased understanding of basic/clinical sciences, strategies that aid individual learning
Study 1016 - not stated
Study 1018 - successful integration of e-learning environments with learning theory strategies

MEDIUM/HIGH
Study 1013 - implied reusability/lecturer LO use/pedagogy course costing
Study 1004 - reduced tutor marking/tutorials/communication with students, increased student discussion
Study 1003 - RLO reusability

MEDIUM/LOW
Study 1030 - increased clinical science understanding
Study 3001 - not stated
Study 1006 - RLO reusability

LOW
Study 1017 - increased pass rate
Studies 1008 - increased pass rate & retention

Generally speaking, studies that had 'strategies to aid learning' as their stated outcome scored high, whilst those that stated 'increased pass rate' scored low – perhaps indicating that making changes to curricula to increase pass rate *alone* is not enough.. Those focusing on reusability appear to score Medium.

APPENDIX 37 - Educational Approach

HIGH
Study 3017 - not stated?
Study 1016 - short course (research)
Study 1018 - learning resource review

MEDIUM/HIGH
Study 1013 - learning/teaching resource review
Study 1004 - cumulative integrated
Study 1003 - learning/teaching resource review/modular non-integrated

MEDIUM/LOW
Study 1030 - free-standing RLO
Study 3001 - modular integrated
Study 1006 - inter-site review?

LOW
Study 1017 - modular non-integrated
Study 1008 - modular non-integrated

Generally speaking the shorter learning resource reviews scored high and modular non-integrated courses scored low.

APPENDIX 38 - Pedagogical Principles

Study 0003: Chickering & Gamson 1989 (Good UG principles of learning)
Components:
Encouraging students/faculty contact
Encouraging co-operation among students
Ancouraging active learning
Giving prompt feedback
Emphasising time on task
Communicating high expectations
Respecting diverse talents and ways of learning
Creators:
Chickering A & Gamson Z, 1989 'Seven principles of good practice in UG education' Racine, W.I: Institutional inventory racine: Johnson Foundation Inc
Chickering AW & Ehrmann S (Oct 1996) Implementing the seven principles: Technology as lever' AAHE Bulletin

Study 0005: Active learning (Hall Hutchings & White 1995/Halliday 1973 & 1975)
Philosophy: Resource-based learning, Tutor makes all pedagogical decisions, Design Action potential networks
1 - Components: 2 layers: Resource later, Use layer
Creators: *Layers*: Hall W, Hutchings G & White S 1995 'Breaking down the barriers: An architecture for developing and delivering resource based learning materials' in J D Tinsley and TJ Van Weert (Eds) World conference on computers in education VI: WCCE '95, Liberating the learner, Chapman and Hall.
Adapted by: T Boyle and J Cook 2002 'Towards a pedagogically sound basis for learning object portability and re-use' Ascilite conference proceedings p101-109
2- Components: Design choices are made on 3 parallel choices made on 3 macro functions: Content structuring (selection and structuring of the learning content), Interactivity (designing for user interaction with this content), Compositional (creation of a coherent composition both within and across contexts)
Creators: *Context grounded systematic networks:* Halliday MAK1973 'The functional analysis of language' Appendix in B Bernstein 'Class codes & control' Vol.3 Routledge and Kegan Hall, AND Halliday MAK1975 'Talking one's way in: A sociolinguistic perspective on language and learning' in A Davies (Ed) 'Problems of language and learning' Heinneman
Adapted by: T Boyle and J Cook 2002 'Towards a pedagogically sound basis for learning object portability and re-use' Ascilite conference proceedings p101-109

APPENDIX 39 - Example of Procedural commentaries

Hughes & Atwell Procedural Commentary

Tell the participant the following:
1. What do you want to be tested on? (Write down what they say)
2. Is there a type of certain evaluation that you prefer? (Write down what they say)
3. Has your experience of learning been a generally positive or negative experience? Why? (Write down what they say)
4. Would you describe your attitude towards learning as generally positive or negative? Why? (Write down what they say)
5. Would you describe your motivation to learn as high or low? Why?(Write down what they say)

DO PACKAGE WITH QUESTIONS INTEGRATED THROUGHOUT (formative evaluation) drawing their attention to the bits they want to be tested on

DO CUMMULATIVE SCENARIO AS TEST (summative evaluation)

1. Do you think you would be able to use packages like this on your own? (Write down what they say)
2. Do you think packages like this would help you to become a more independent learner i.e. needing less help from colleagues, courses, etc? (Write down what they say)

DO USABILITY QUESTIONNAIRE

APPENDIX 40 - The Whole Grounded Theory

This was plotted on a very large piece of paper and causative arrows assigned. Every single verbatim participant response was followed along the arrows to ensure that every single phrase fitted exactly into the theory. This is too large to include here but is available on request from the author.

APPENDIX 41 - IHEP Distance Learning Benchmark Categories

The table below shows IHEP Distance Learning Benchmark Categories:

Category	Description
Institutional Support	Activities by the institution that help to ensure an environment conducive to maintaining quality distance education, as well as policies that encourage the development of Internet-based teaching and learning including technological infrastructure issues, a technology plan, and professional incentives for faculty.
Course Development	The development of courseware, which is produced largely either by individual faculty (or groups of faculty members) on campus, subject experts in organizations, and/or commercial enterprises
Teaching Learning Process	Activities related to pedagogy including interactivity, collaboration, and modular learning.
Course Structure	Policies and procedures that support and relate to the teaching/learning process, including course objectives, availability of library resources, types of materials provided to students, response time to students, and student expectations.
Student Support	Student services normally found on a college campus including student training and assistance while using the Internet.
Faculty Support	Activities that assist faculty in teaching online, including policies for faculty transition help as well as continuing assistance throughout the teaching period.
Evaluation and Assessment	Policies and procedures that address how, or if, the institution evaluates Internet-based distance learning including outcomes assessment and data collection.

From Jurczyk et al 1999

APPENDIX 42 - Unique Questions Added For Mobile Learning

The unique questions to the generic protocol for mobile learning data extraction were the classification of software, learning theories used, type of ISD, pedagogical frameworks/pardigms, evaluation framework capabilities, specific features of evaluations e.g. its comstituent parts, advantages, and disadvantages. (This is available on request from the author).
