AUGMENTED CONVERSATION AND COGNITIVE APPRENTICESHIP METAMODEL BASED INTELLIGENT LEARNING ACTIVITY BUILDER SYSTEM

THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF DE MONTFORT UNIVERSITY, UNITED KINGDOM, FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

By

ADETOKUNBO ABAYOMI ABDURRASHEED ADENOWO

DECEMBER 2012

Abstract

This research focused on a formal (theory based) approach to designing Intelligent Tutoring System (ITS) authoring tool involving two specific conventional pedagogical theories—Conversation Theory (CT) and Cognitive Apprenticeship (CA). The research conceptualised an Augmented Conversation and Cognitive Apprenticeship Metamodel (ACCAM) based on apriori theoretical knowledge and assumptions of its underlying theories. ACCAM was implemented in an Intelligent Learning Activity Builder System (ILABS)—an ITS authoring tool. ACCAM's implementation aims to facilitate formally designed tutoring systems, hence, ILABS—the practical implementation of ACCAM—constructs metamodels for Intelligent Learning Activity Tools (ILATs) in a numerical problem-solving context (focusing on the construction of procedural knowledge in applied numerical disciplines). Also, an Intelligent Learning Activity Management System (ILAMS), although not the focus of this research, was developed as a launchpad for ILATs constructed and to administer learning activities. Hence, ACCAM and ILABS constitute the conceptual and practical contributions that respectively flow from this research.

ACCAM's implementation was tested through the evaluation of ILABS and ILATs within an applied numerical domain—the accounting domain. The evaluation focused on the key constructs of ACCAM—cognitive visibility and conversation, implemented through a tutoring strategy employing Process Monitoring (PM). PM augments conversation within a cognitive apprenticeship framework; it aims to improve the visibility of the cognitive process of a learner and infers intelligence in tutoring systems. PM was implemented via an interface that attempts to bring learner's thought process to the surface. This approach contrasted with previous studies that adopted standard Artificial Intelligence (AI) based inference techniques. The interface-based PM extends the existing CT and CA work. The strategy (i.e. interface-based PM) makes available a new tutoring approach that aimed fine-grain (or step-wise) feedbacks, unlike the goal-oriented feedbacks of model-tracing. The impact of PM—as a preventive strategy (or intervention) and to aid diagnosis of learners' cognitive process—was investigated in relation to other constructs from the literature (such as detection of misconception, feedback generation and perceived learning effectiveness). Thus, the conceptualisation and implementation of PM via an interface also contributes to knowledge and practice.

The evaluation of the ACCAM-based design approach and investigation of the above mentioned constructs were undertaken through users' reaction/perception to ILABS and ILAT. This involved, principally, quantitative approach. However, a qualitative approach was also utilised to gain deeper insight. Findings from the evaluation supports the formal (theory based) design approach—the design of ILABS through interaction with ACCAM. Empirical data revealed the presence of conversation and cognitive visibility constructs in ILATs, which were determined through its behaviour during the learning process. This research identified some other theoretical elements (e.g. motivation, reflection, remediation, evaluation, etc.) that possibly play out in a learning process. This clarifies key conceptual variables that should be considered when constructing tutoring systems for applied numerical disciplines (e.g. accounting, engineering). Also, the research revealed that PM enhances the detection of a learner's misconception and feedback generation. Nevertheless, qualitative data revealed that frequent feedbacks due to the implementation of PM could be obstructive to thought process at advance stage of learning. Thus, PM implementations should also include delayed diagnosis, especially for advance learners who prefer to have it on request. Despite that, current implementation allows users to turn PM off, thereby using alternative learning route. Overall, the research revealed that the implementation of interface-based PM (i.e. conversation and cognitive visibility) improved the visibility of learner's cognitive process, and this in turn enhanced learning—as perceived.

Acknowledgement

This research could not have been undertaken without the significant role played by some individuals. Their unflinching support sees me through the long journey. In this regard, I sincerely express my profound appreciation to my supervisors, Ashok Patel and Dr. Richard Hall, for their all-round support and positive critique. I recognise their interests in my well-being, and appreciate their efforts in boosting my morale at times when the spirit was low. Essentially, I must recognise the very significant role Ashok Patel played in getting my visa through and all forms of assistance he rendered at times of difficulty. This is a very rare supervisor-student relationship. I must equally recognise the contributions of the Head of Department, Dr David Russell, who participated in the interview process and his words of encouragement, the very first time I met him on arrival at the university. I also recognise the very important roles played by Professor Elaine Harris and Professor Martyn Denscombe during my studies.

My profound appreciation goes to all my friends and colleagues for their memorable and generous contributions. I have chosen not to mention names, realising the long list and do not want to hurt anyone. Yet, I must sincerely express that I am greatly indebted to you all.

This thesis won't be complete without recognising my siblings and in-laws for their support. More importantly, I express my heartfelt gratitude to my immediate family for their patience, endurance and love. Without it, this journey couldn't have been embarked upon, let alone seeing it through. So, thank you dear beloved ones.

Also to the Departmental Secretary, Faculty IT Technicians, Graduate School Office staff, Security personnel and Student Housing/Accommodation staff, I say: thank you all for providing support and help during the research period—a beautiful/unique experience in life.

Lastly, I wish to acknowledge that this research was part-supported by De Montfort University.

Contents

Abbreviations

Acronym	Meaning
ACT-R	Adaptive Control of Thought Rational
ACCAM	Augmented Conversation and Cognitive Apprenticeship Metamodel
AI	Artificial Intelligence
ANOVA	Analysis Of Variance
ASPIRE	Authoring Software Platform for Intelligent Resources in Education
BDASUM	Builder Assumption scale
BDRST	Builder Restriction scale
BIP	Builder Interview Protocol
BQ	Builder Questionnaire
CA	Cognitive Apprenticeship
CAI	Computer Assisted Instruction or Computer Aided Instruction
CAL	Computer Aided Learning
CASTE	Course Assembly System and Tutorial Environment
CBM	Constraint Based Modeling
CBT	Computer Based Training
CCVSB	Conversattion aid Cognitive Visibility scale
CI	Confidence Interval
CPVSB	Cognitive Process Visibility scale
CREAM-Tools	Curriculum Representation and Acquisition Model-Tools
CT	Conversation Theory
CTAT	Cognitive Tutor Authoring Tools
CVSBLN	Cognitive Visibility and Learning scale
DEPTHS	DEsign Pattern Teaching Help System
EPIC	Executive Process Interactive Control
eTQ	eTutor Questionnaire
FBM	Feature Based Modeling
GBLE	Groupware-Based Learning Environment
GCE	General Certificate of Education
GCSE	General Certificate of Secondary Education
HCI	Human Computer Interaction
IBM	International Business Machine
ILABS	Intelligent Learning Activity Builder System
ILAMS	Intelligent Learning Activity Management System
ILAT	Intelligent Learning Activity Tool
ITS	Intelligent Tutoring System
ITT	Intelligent Tutoring Tool
KR	Knowledge Representation
LAMS	Learning Activity Management System
LEMONADE	Learning Environment for MObile Network-Able Devices

LNEFTV	Learning Effectivenes scale
M	Mean
MELCOE	Macquarie University E-learning Centre of Excellence
M-individuals	Machines or computer processors
MISCP	Misconception scale
NECO	National Examinations Council
PAT	Pump Algebra Tutor
PDTIM	Production Time scale
PE	Performance Error
P-individuals	Psychological individuals
PLATO	Programmed Logic for Automatic Teaching Operations
PROUST	PROgram Understand for Students
QUIS	Questionnaire for User Interaction Satisfaction
REDEEM	Reusable Educational Design Environment and Engineering Methodology
RELFDBK	Relevant Feedback scale
Rho	Spearman Correlation
SD	Standard Deviation
SOAR	State-of-the Art computational theory or States, Operators And Reasoning architecture
SPSKL	Special Skill scale
TIMFDBK	Timely Feedback scale
TLTP	Teaching and Learning Technology Programme
TUTBHV	generated Tutor Behaviour scale
TUTSTRG	Tutoring Strategy scale
UNCLE	Using Notes for a Case-based Learning Environment
USAB	Usability construct
WAEC	West Africa Education Council
WEAR	WEb based authoring tool for Algebra Related domains
XAIDA	eXperimental Advanced Design Advisor
xPST	Extensible Problem Specific Tutor
ZPD	Zone of Proximal Development

Tables

Figures

Figures in Appendix:

Chapter 1: Introduction

This chapter presents a summary account of the issues that triggered the current investigation and clarifies its aim/objectives. The scope and significance of this research are also stated. Finally, an outline of the entire thesis is presented.

1.1 Background

Intelligent Tutoring System (ITS) and ITS authoring can be classified under the umbrella of Technology Enhanced Learning (TEL). ITS is a specific type of tutoring systems which includes Computer Assisted Instruction (CAI) and Computer Supported Collaborative Learning (CSCL). ITS provides an intelligent and/or adaptive teaching and learning environment for individualised learning (Koedinger & Corbett, 2006; Ben Ammar et al., 2010) unlike CAI systems, that are non-adaptive individualised learning environments; on the other hand, CSCL is group-based teaching-learning environment (Hartley, 2010). TEL encompasses tutoring systems, authoring systems and Learning Management Systems (LMS)―a system that enables the administration of teaching and learning activities. TEL enables individual and/or collaborative teaching-learning through diverse technologies, such as a stand-alone computer, web 2.0 and mobile communications. Thus, ITS is a typical instance of TEL due to the provision of individualised learning, while ITS authoring system is a framework—sometimes including an interface—that enables the contruction of ITSs (Murray, 2003a).

ITS evolved from CAI or, as they are sometimes called, Computer Aided Learning (CAL) systems (Freedman, 2000; Keles et al., 2009). ITS is distinguished from CAI in that it adapts to users' individual needs (Martin, 2001; Siddappa, Manjunath & Kurian, 2009; Zarandi, Khademian & Minaei-Bidgoli, 2012). Adaptation was achieved by modelling learners' behaviour and adjusting tutoring strategy during learning of the target domain knowledge (Martin, 2001). The domain model, tutor model and student model constituted three out of four commonly presented components of an ITS in the literature. The fourth was refered to as the interface—or communication—component (Alves, Pires & Amaral, 2009; Woolf, 2009). These four components communicated based on the learner's interaction with the tutoring system in order to provide courseware, feedback/hints and problem generation/selection (Jackson, 2002). Also, an

ITS provided a scaffolding environment that enabled the learner to practise skills within a target domain (Martin, 2001). Scaffolding is the step-by-step support of learning in such a way that guidance and/or learning tasks are provided to a learner during problem-solving based on his/her needs (Aleven et al., 2009).

ITS was inspired, firstly, by the need to model human "intelligence" in a technology-driven tutoring system, dating back to works in the 1970s (Siddappa, Manjunath & Kurian, 2009; Li, Zhuying & Bing, 2010), and the conception of ITS in the 1980s (Pozzebon et al., 2007). Secondly, it derived further inspiration from the empirical studies on human tutors (Lane & Johnson, 2008), which informed attempts to mimic human one-to-one tutoring strategy in technology-driven tutoring systems (Graesser, Conley & Olney, 2012). Bloom's (1984) finding that human one-to-one tutoring was more effective, in contrast to one-to-many, provided significant encouragement to the ITS approach to learning. While the above highlighted inspirational factors contributes to extensive research in the ITS field—as elaborated later below, its authoring research has been relatively less explored. The later research field focuses on how intelligent tutors (i.e. ITSs) can be produced quickly and relatively easily by human tutors who might not possess programming skills (Hsieh & Hsieh, 2001; Murray, 2003a; Ainsworth & Fleming, 2006; Talhi et al., 2007; Blessing et al., 2009; Direne et al., 2009; Mitrovic, Martin & Suraweera, 2009; Gilbert et al., 2011).

Since the inception of these two research areas (i.e. ITS and ITS authoring), research efforts have grown in order to provide more useful and effective ITSs for classroom usage and research purposes. ITS implementations resulting from these efforts utilised several techniques, including Artificial Intelligence (AI) methodologies. Some of the AI techniques explored included fuzzy logic and Bayesian Networks (BNs) (Conati, Gertner & VanLehn, 2002; Pena, Sossa & Gutierrez, 2007; Conati, 2009; Chieu et al., 2010; Zarandi, Khademian & Minaei-Bidgoli, 2012), while some explored Knowledge Representation (KR) schemes (Hatzilygeroudis & Prentzas, 2004). Also, several approaches had been implemented in the ITS domain and student models. Mitrovic (2012) acknowledged two major approaches, namely the Model/Knowledge tracing approach—promoted by researchers at Carnegie Mellon University (Anderson, Boyle & Yost, 1985; Koedinger et al., 1997) and the Constraint Based Modelling (CBM) approach proposed by Ohlsson (1992). Model tracing is an approach that compares a learner's outcome or solution goal/sub-goal with the expert model of an ITS; thus, it tracks a learner's progress by generating solutions step-by-step (Mitrovic & Weerasinghe, 2009). On the other hand, CBM is the conception/representation of knowledge within an ITS in the form of constraints. Constraints represent a set of syntax and semantics on correct solutions of target domain knowledge (ibid.).

Past studies have considered issues on adaptation, personalisation (Brusilovsky, 1999; Phobun & Vicheanpanya, 2010), development of stand-alone and web-capable intelligent tutors (Keles et al., 2009; Quinton, 2010). In some cases, cognitive issues were explored (Anderson et al., 1984, 1987; Aleven et al., 2004, 2006c; Muldner & Conati, 2010). In rare cases, links between educational theories and ITS development were claimed (Conati & VanLehn, 1996; Patel, Scott & Kinshuk, 2001; Siang & Rao, 2003; Lee, Cho & Choi, 2006). Significantly, informal theories were cited, as highlighted by Dessus, Mandin & Zampa (2008), and Keles et al. (2009), while some studies had no recourse to any theory since they cannot be traced.

In terms of domains covered by past ITS research efforts, mathematics (including algebra, geometry and numerical methods), physics, programming and engineering, enjoyed reasonable patronage (Hsieh & Hsieh, 2001; Matsuda & VanLehn, 2005; VanLehn et al., 2005; Sykes, 2005; Siddappa, Manjunath & Kurian, 2009; Arroyo, Royer & Woolf, 2011; Cheema & LaViola, 2012). In contrast, applied numerical domains, such as accounting and finance, witnessed very little effort. Byzantium Intelligent Tutoring Tools (ITT) happened to be a pioneer work (Kinshuk, Patel & Russell, 2000; Patel, Cook & Spencer, 2003) in the latter domain. The Byzantium ITT has been used extensively in classrooms, and has stood the test of time in some universities in the United Kingdom (Stoner, 2003; Patel, Cook & Spencer, 2003). Thus, extending the research in the applied numerical domain, to explore how to produce more useful and reliable ITSs, was considered a research window.

Despite the above efforts, only a few reliable ITSs can be found in classrooms. Several factors have been attributed which revolve round its nature and development. These included the facts that building intelligent tutors required complex reasoning, and was noted to be difficult, expensive, time-consuming and requiring the collaboration of experts in related fields (Woolf & Cunningham, 1987; Virvou & Moundridou, 2001; Moundridou & Virvou, 2003a; Murray, 2003a; Martin, Mitrovic & Suraweera, 2008; Woolf, 2009; Suraweera, Mitrovic & Martin, 2010; Zarandi, Khademian & Minaei-Bidgoli, 2012). These factors limited the number of useful ITSs that could be constructed, thereby imposed a major bottleneck in their use (Murray, 1997). In order to avert/reduce the highlighted and other related problems, and enhance ITS construction in a cost-effective manner, ITS authoring emerged as an explorable research field.

ITS authoring research efforts yielded some reasonable results. Among such fruitful efforts are: **Eon**—is the name for a suite of domain-independent tools for authoring all aspects of a knowledge-based tutor (Murray, 1998, 2003b); **R**eusable **E**ducational **D**esign **E**nvironment and **E**ngineering **M**ethodology (REDEEM)—(Major, Ainsworth & Wood, 1997; Ainsworth et al., 2003), **IRIS—**an authoring tool that derived its pedagogical requisites from a cognitive theory of instruction and developed to build intelligent tutoring systems in a variety of domains (Arruarte et al., 2003); **C**urriculum **RE**presentation and **A**cquisition **M**odel **Tools** (CREAM-Tools)—an authoring environment for curriculum and course building in ITS (Nkambou et al., 2003), **We**b-based authoring for **A**lgebra **R**elated-domains (WEAR)—(Moundridou & Virvou, 2001a; Virvou & Moundridou, 2000); **C**ognitive **T**utor **A**uthoring **T**ools (CTAT)—an authoring tool for creating cognitive (i.e. rule-based) ITSs and example-tracing ITSs (Aleven et al., 2006a; Aleven et al., 2006b; Blessing et al., 2009); **A**uthoring **S**oftware **P**latform for **I**ntelligent **R**esources in **E**ducation (ASPIRE)—an authoring and deployment environment for constraint-based intelligent tutoring systems (Mitrovic, Martin & Suraweera, 2009); and **L**earning **E**nvironment for **MO**bile **N**etwork-**A**ble **DE**vices (LEMONADE)—a framework for planning and conducting field trips with mobile devices (Giemza et al., 2010).

The above identified tools were designed to generate ITSs within a short space of time and to eliminate the developmental expertise required. They aimed to reduce the cost and time associated with the development of individual ITSs. Each of the above ITS authoring tools employed different strategies, approaches and mechanisms to achieve their design goals. However, many of these tools were also not formalised using educational theories. The foregoing tended to limit their educational value because their development was more or less driven by AI and/or cognitive models/techniques or human computer interaction standards rather than pedagogical engineering. Thus, the foregoing constituted the foundation of the research discussed in this thesis.

1.2 The Research Motivation

The implementation of the human one-to-one tutoring strategy—which has been proven to be an effective strategy in Bloom's (1984) studies as cited in Koedinger & Corbett (2006) and Woolf (2009)—in ITSs and the lack of a formal link between ITS design and theory (as noted by Self, 1990b), raises the need to formalise ITS design. Moreover, VanLehn, Jones & Chi (1992, p.54) claimed that "*a good theory of knowledge acquisition methods could improve the design of instructional situations.*" Also, the interplay between theory and practice has been established for ITS-precursors—that is, CAI systems (Koschmann, 2001; Hartley, 2010). Notionally, educational theories should shape ITS design/development since it has been noted that this technology was rooted in the human one-to-one tutoring approach—an educational tutoring strategy. On that note, this research sees the need to establish apriori link between conventional pedagogical theories and ITS construction.

In search of appropriate pedagogical theories, this research acknowledged certain pedagogic activities that take place in a conventional teaching environment. A human tutor engages the learner in conversation to achieve learning. Conversation—a verbal and/or non-verbal information exchange between two or more cognitive systems (i.e. individuals and/or computer processors) (Klemm, 2002; Holland & Childress, 2008)—is a concept embraced in Gordon Pask's Conversation Theory (CT) (Pask, 1976a, 1976b, 1976c, 1988; Scott, 2001a; Scott, 2001b) and reviewed/applied in other works (Boyd, 2004; Sharples, 2005; Heinze, Procter & Scott, 2007; Heinze & Heinze, 2009; Scott & Cong, 2010). Cognitive Apprenticeship (CA) (Collins, Brown & Newman, 1989) provides a framework that supports learning domains that involve a lot of cognitive tasks through engagement in situated activities. By situated activities, the thesis implies the framing of learning activities in a way that matches real world situations or practice. Moreover, Collins, Brown & Holum (1991) argued that the human tutor can provide useful guidance if the learner's cognitive or thinking process can be made visible. Thus, augmenting conversation within a cognitive apprenticeship framework could be explore, since neither CT nor CA explicitly provided means of achieving improved cognitive visibility. Augmentation of conversation aimed to monitor a learner's thought process while engaging in information exchange with a tutoring system via an interface (i.e. visual calculator). This is intended to improve the visibility of a learner's thought process, regarded as cognitive visibility in this thesis, while the practical implementation of cognitive visibility is referred to as Process Monitoring in this work.

Also, previous studies indicated that students learn from performance error—a notion put forward by Ohlsson (1996b). This translated into Constraint Based Modelling (CBM), as noted by Mitrovic (2012). CBM underpins several studies including works by Martin, Mitrovic & Suraweera (2008), Mitrovic, Martin & Suraweera (2009), and Suraweera, Mitrovic & Martin (2010). CBM tends to capture the learner's knowledge states in the form of constraints and provides feedback according to the states. Furthermore, Ohlsson (1996b) noted that learning from error involves two cognitive functions, namely detecting and correcting error. The error states of a learner must be identified in order to provide corrective feedback. The former could be referred to as detecting misconception or the misconception state, while the latter could refer to corrective feedback. However, a feedback does not have to be corrective only (i.e. negative feedback), since the human tutor being mimicked in the ITS, also provides positive feedback that confirms the learner's correct actions (Mitrovic, 2012). This brings to the fore the concepts of timing and relevance of feedback in relation to the learner's knowledge states. Although, addressing the error states (or misconception)—through provision of timely and relevant feedback—enhances learners ability to make progress (Melis, 2005; Barnes & Stamper, 2010; Stamper, Barnes & Croy, 2011; Mitrovic, 2012). Confirming a learner's correct action could also deepen understanding of the target domain and improve the learning rate (Ohlsson et al., 2007; Barrow et al., 2008; Cade et al., 2008; Di Eugenio et al., 2009; Boyer et al., 2008, 2011; Mitrovic, 2012).

In the light of the above—that is, formalisation of ITS design and consideration of the two cognitive functions—the twin application of earlier-mentioned theories in ITS raises some questions. Is it possible to make a cognitive process visible through conversation? Can cognitive visibility enable detection of learner misconception or missing conception? Does cognitive visibility enable the generation of relevant feedback in response to misconceptions or missing conception? Does cognitive visibility enhance learning effectiveness? Addressing these questions in this research is considered vital, since feedback—along with detection of misconception—is considered a key success factor for an ITS (Shute, 2008). Although some ITS research (e.g. Melis, 2005; Zakharov, Mitrovic & Ohlsson, 2005; Vasilyeva et al., 2007; Ferreira & Atkinson, 2009; Scheuer et al., 2010) has examined feedback, no previous ITS studies have addressed them (i.e. feedback and detection of misconception) especially, through a formal (theory-based) approach using a pedagogic metamodel. The latter referred pedagogic metamodel is a conceptual model—conceived as an augmented conversation within CA framework—that describes the knowledge/representation of the models (i.e. domain model, tutoring model, student model and interface model) constituting an ITS (see chapter 3, section 3.4 for details). Thus, the above questions could be examined through process monitoring (PM) that augments conversation in a cognitive apprenticeship framework when implemented in an ITS. This contrasts with previous CT- and/or CA-based ITS studies that did not implement PM (e.g. Patel, Scott & Kinshuk, 2001; Cameron, 2009). Once again, note that PM was an interface-based tracking of the cognitive activities of a learner and conceptually known as cognitive visibility in this thesis.

In addition, though some successes have been attained regarding ITS authoring, as indicated above, there are still unexplored gaps and some were identified in Murray (1999; 2003a). However, none has been designed for and/or extensively evaluated in the problem-solving context of the applied numerical domains—such as accounting and finance. This gap is being considered in this research, since a key factor that could contribute to successful implementation of ITS authoring tools is to limit them to particular domains or knowledge types (Virvour & Moundridou, 2000; Murray, 2003a). As well, one cannot guarantee existing ITS authoring tools being applicable to generate meaningful intelligent tutors in the problem-solving context of the applied numerical domain—for example, the accounting and finance, which is the evaluation context/domain of this research.

Also, if ITS development is formalised as stated above, ITS authoring tools too should be underpinned by educational theories, but this has not been the trend. An exception is Hayashi, Bourdeau & Mizoguchi (2009), a work acknowledged in Nkambou, Bourdeau & Mizoguchi (2010) as a step towards formalisation of ITS authoring tools using pedagogical theories and response to Self's (1990b) call (the formalisation of ITS design). Nkambou, Bourdeau & Mizoguchi's (2010) acknowledgement further confirms the near absence of pedagogy theory-based formalisation in the ITS/Authoring literature. However, the study (Hayashi, Bourdeau & Mizoguchi, 2009) utilises multiple theories using what they called ontological engineering approach. While reviewing ITS authoring tools, Hayashi, Bourdeau & Mizoguchi (ibid) acknowledged that their SMARTIES authoring tool was too complex for non-programmers. It appears to lack pedagogical focus due to its many theories and complexity. This threatens its eventual utilisation by curriculum designers (mostly non-programmers), who might want to purchase tools that address specific pedagogical needs, as argued by Murray (2003a). Moreover, SMARTIES real world utilisation cannot be guaranteed since it has not been proven. In addition, it falls within the class of pedagogy-oriented tools, thus lying outside the ambit of this research, which focuses on the performance-oriented category, when considered, based on Murray's (1999, 2003a) categorisation of authoring tools. Therefore, research efforts channelled towards theory-based formalisation of ITS authoring tools' design, which are without the engineering complexity of Hayashi, Bourdeau & Mizoguchi's (2009) approach, are still required. Moreover, ITS authoring research will be uniquely placed in the literature if driven by educational theories, since none exist for now in the performance-oriented category of applied numerical domains.

Additionally, an ITS authoring tool design should satisfy a key factor that contributed to the success of ITS for it to be useful. This is the construction of ITSs that provide timely and relevant guidance/feedback, which is deemed significant in pedagogy (Melis, 2005; Shute, 2008). Thus, an ITS authoring tool designed based on a formal link to pedagogy theories would be appropriate. This equally raises some questions: how can we achieve formalisation of such an ITS authoring tool design in order to construct ITSs that provide effective feedback in a conversation and cognitive visibility learning environment? Also, how do we ensure that such a formalised ITS authoring tool is usable by authors, who are non-programmers, for it to achieve its purpose? Answers to these questions could also enhance the investigation of the research issues raised with respect to ITS formalisation (as stated above). Thus, this research aims to address these latter questions along with earlier raised ones.

1.3 Research Aim / Objectives

In the light of the above, this research addresses the formalised construction of ITSs to alleviate or eliminate difficulties confronting students acquiring procedural knowledge in the problem-solving context of applied numerical disciplines. Procedural knowledge is the practical skill required to solve problems in a target domain, while declarative knowledge—the other type of knowledge—is the abstract or conceptual knowledge which may include target domain concepts, terms and their relationships that is acquireable through reading and/or instructions (Ohlsson & Mitrovic, 2006; Akin, 2008). The formalised construction of ITS was actualised through apriori link between pedagogy theories and an ITS authoring tool. The apriori link was undertaken through the conception and implementation of a pedagogic metamodel in an ITS authoring tool, while the metamodel involved the augmentation of learning conversations within a cognitive apprenticeship framework. Note that the augmentation of learning conversations was regarded as PM in this work. In this thesis, ITS authoring tool was also referred to as an Intelligent Learning Activity Builder System (ILABS), while ITS was equally known as Intelligent Learning Activity Tool (ILAT).

Based on this aim, the following four research objectives were identified:

- To conceptualise and test the implementation of a pedagogic metamodel in ILABS.
- To assess the usability of the implemented metamodel (i.e. ILABS).
- To evaluate the use of process monitoring to increase visibility of the cognitive process of a learner.
- To determine the perception of target users regarding the impact of process monitoring on feedback and learning effectiveness.

1.4 Research Questions

In order to achieve the above stated research objectives, two research questions were developed. Each question was further broken down into four propositions as stated below.

Question One: What is the perception of users on the conception and implementation of a pedagogic metamodel in ILABS, which can be utilised by non-programmers to generate an unrestricted number of intelligent tutoring systems in a numerical problem solving context of applied numerical domains?

- ***Proposition 1.1:*** A pedagogic metamodel can be conceptualised and implemented in ILABS for the applied numerical problem solving context.
- ***Proposition 1.2:*** It is possible to generate tutoring systems that support process monitoring and model-tracing from the implemented metamodel (i.e. ILABS).
- ***Proposition 1.3:*** The implemented metamodel can be used to create an unrestricted number of tutoring systems within a short space of time by authors (i.e. lecturers) who are non-programmers.

- ***Proposition 1.4:*** Users of the implemented metamodel have a positive perception about its ease of use and usability.

Question Two: Can the learner's cognitive process be made visible to aid the generation of relevant and timely diagnostic feedback in order to enhance learning effectiveness in the numerical problem solving context?

- ***Proposition 2.1:*** The learner's cognitive process can be made visible to the tutoring system (or domain expert).
- ***Proposition 2.2:*** Cognitive visibility can be used to aid the generation of relevant and timely diagnostic feedback.
- ***Proposition 2.3.*** Cognitive visibility exposes/tracks learner's misconceptions.
- ***Proposition 2.4:*** Cognitive visibility enhances learning effectiveness.

1.5 Scope of the Research

This applied research involved the conceptualisation of a pedagogic metamodel (ACCAM—Augmented Conversation and Cognitive Apprenticeship Metamodel) and subsequent implementation in ILABS. The research developed, at a small scale level, an Intelligent Learning Activity Management System (ILAMS)—a platform used to administer and launch the ILATs contructed through the ILABS (see chapter 3 for details), although it was not the main focus of this research. Also, the research included the evaluation of the ILABS and its products (i.e. ILATs or ITSs generated from it). The evaluation was designed to be a multi-institutional evaluation involving at least two institutions. It was principally a quantitative research, but it additionally employs qualitative approach to derive deeper insight.

Participants were lecturers (the target users of the ILABS and herein referred to as authors) and students (the envisaged users of the ILATs and herein referred to as learners). With the research design employed, a large data set was collected and analysis undertaken. The feasibility of the formal (theory-based) approach adopted in this work was confirmed, especially as applicable to the numerical problem-solving context of applied numerical domains—using accounting as the evaluation domain. It also enabled the confirmation/refutation of the theoretical assumptions that underlies the pedagogic metamodel utilised in the research.

1.6 Significance of the Research

The research presented in this thesis extended the Byzantium project that was earlier mentioned. The research also looked beyond the Byzantium project and learned from other works in the ITS/authoring literature which were referenced throughout this thesis. This research approach was upheld in order to: [i] identify the examinable research issues in the ITS Authoring research field; and [ii] enable this research to contribute to knowledge and practice.

Thus, with the adoption of a formal (theory-based) approach—utilising a pedagogic metamodel design that is underpinned by educational theories—and the investigation of PM—as a learning intervention and cognitive process diagnosis enhancer, this work intended to be a unique research endeavour in that:

i. It provided a better understanding of how theory can translate into practice, since—at least—it afforded the opportunity to conceptualise and test the implementation of ACCAM in ILABS—that was utilised to generate tutoring systems in the applied numerical domains, thus answering the call of Self (1990b) that ITS design should be formalised. Also, the success of the approach either confirmed or refuted the argument in the literature (Murray, 1999; Virvou & Moundridou, 2000), namely to produce powerful and usable intelligent tutoring systems, an ITS authoring tool should be limited to particular domains or knowledge types; and that an ITS authoring tool for every possible domain is not feasible.
ii. No such work currently exists—involving an apriori link between a pedagogic metamodel and an ITS authoring tool—that has been extensively evaluated in an applied numerical discipline, e.g. accounting, as far as we know.
iii. It prepared the ground for the extension of this approach to other numerical domains.
iv. It demonstrated an alternative approach for achieving intelligence in tutoring systems (without using the standard AI techniques, e.g. BNs, fuzzy logic etc.) and improving cognitive visibility (to aid detection of misconception and enhance feedback generation/learning effectiveness), through the implementation of PM via an interface—that brings learner's thought process to the surface.

Also, the theoretical foundation of this research enabled the exploration of the underlying assumptions of the theories considered (CT and CA). The assumptions, as conceived in the pedagogic metamodel conceptualised, are intended to promote learners' engagement in conversation that could aid cognitive process visibility through process monitoring, then utilised process monitoring to guide the generation of feedback/hints during learning. Although research concepts related to cognitive visibility—such as plan recognition, cognitive mapping—have been undertaken in the past, further investigation of cognitive visibility in this thesis was undertaken through a pedagogic metamodel approach—based on two theories—and took an ITS authoring route, unlike previous works that utilised AI and/or CA in ITSs only. Also, its conceptualisation and implementation was clearly different from other related concepts, since it relied on an interface managed through a generic algorithm developed in this work. This introduces a unique way of achieving cognitive visibility, the generation of feedback/hints in an ITS, and enhanced the investigation of the conceptual link between cognitive visibility and feedback/misconception. Therefore, this work stands out when compared to approaches implemented in relatively close research concepts reported in the literature.

Furthermore, PM—used as a preventive (i.e. interventionist) and diagnostic strategy—was implemented as a feature provided by an ITS authoring tool which is explicitly linked to theories that underpinned above referred pedagogic metamodel, CT (Pask, 1976a, 1976c; Scott, 2001a; Heinze & Heinze, 2009) and CA (Collins, Brown & Newman, 1989; Collins, 1991a; Collins, Brown & Holum, 1991; Denne & Burner, 2008). Also, the ITS authoring tool developed using the pedagogic metamodel, has the capacity to generate intelligent tutors with dual tutoring strategies—one through model-tracing (already used in Byzantium) and interface-based PM (newly-introduced). Note that, unlike PM (the step-wise or sub-goal monitoring of learner's thought process), model tracing is the comparison of a learner's outcome or solution goal with an expert model of an ITS as earlier defined. Authors (i.e. lecturers) could—during authoring—determine whether to include dual-tutoring strategies or one of the available strategies when building ITSs for learners (i.e. students). It would also afford flexibility for learners, in that, he/she could choose, during learning, any of the available tutoring strategies embedded in the ITS constructed.

1.7 Summary

In all, this thesis contains seven chapters inclusive of the current chapter, which provides an introduction to the research and a summary of what was covered in this work. Chapter 2 covers the review of the ITS/Authoring literature and establishes the theoretical frameworks for the research. In chapter 3, the theoretical frameworks are considered in detail leading to the conception of a pedagogic metamodel that was implemented in an ITS authoring tool. Chapter 4 treats the methodological approach undertaken to evaluate the implemented metamodel and associated products, and includes the justification for the chosen approach. Chapters 5 and 6 provide the analysis of data collected with respect to research questions one and two respectively; they also include a discussion of findings with respect to previous works in the field. On a final note, chapter 7 contains the conclusion to the current work, stating the key findings, contributions to knowledge and practice in the ITS authoring field, recommendations, and identifies other areas that can be addressed in future research.

Chapter 2: Tutoring Systems—The Intelligent Learning Approach

"Complex ideas are built from simple ideas that are gathered from the world around him." by Jean-Jacques Rousseau

(excerpt from Hilgard & Bower, 1975)

This chapter provides background information on the tutoring systems field, to aid comprehension of research presented in this thesis. Generally, tutoring systems are technology-based teaching and learning environments. However, a review of the literature with respect to a specific type of tutoring systems, the Intelligent Tutoring System (ITS), and related authoring research was undertaken. This enhanced the emergence of the research issues addressed in this thesis. Relevant theoretical frameworks were identified. These frameworks underpinned the conception and implementation of a pedagogic metamodel in an "*Intelligent Learning Activity Builder System*" (ILABS) which was meant to construct an inventory of "*Intelligent Learning Activity Tools*" (ILAT). On the other hand, the actual design, implementation and evaluation of ILABS and ILAT were discussed in subsequent chapters.

In this research, the use of ILABS and ILAT assumes the full meanings associated with their alternate terms/phrases in the literature, that is, "ITS authoring tool" and "ITS" (also known as "Intelligent Tutor") respectively. These phrases are interchangeably used and imply the same meaning. They assume the theoretical underpinnings of contributing disciplines to the field, depending on the issue(s) being investigated. Note that research issues relating to ITS and ITS authoring tools are discussed in detail in this chapter. A general view of ITS/Authoring research is provided, but at certain points, the research is contextualised within the numerical problem solving context of applied numerical disciplines, and concretises by implementing/evaluating within the numerical aspects of the accounting domain. This became necessary since research was undertaken in context, and each discipline works within its own frame of reference (Luckin, 2010). Moreso, contextualising—while still attempting to generalise—was considered appropriate (Nardi, 1996 cited in Luckin, 2010).

In order to achieve the above, section 2.1 addresses the general field of tutoring sytems and demonstrates the emergence of ITS research. Section 2.2 further surveys the ITS field in fine detail, the platform that informed the conception and implementation of a metamodel in an ITS authoring tool. Section 2.3 discusses the theoretical contribution of three main disciplines that impacted ITS research since inception. Section 2.4 considers existing approaches to ITS implementation, while section 2.5 provides insights into the components that constitute ITS architecture. Some selected examples of intelligent tutors in the numerical domain are reviewed in section 2.6. In section 2.7, a review of authoring research is presented, followed by section 2.8 that provides an outlook of what this research intends to address. Section 2.9 briefly discusses the theoretical frameworks which underpin the conception and implementation of a pedagogic metamodel (see details in chapter 3). Section 2.10 provides the research questions/propositions addressed. Finally, the chapter concludes with a summary in section 2.11.

2.1 Teaching and Learning Systems—A General Outlook

Dynamics in education yielded revision and re-evaluation of traditional teaching and learning techniques, and the introduction of new approaches to pedagogy, a continuous trend as education faces new challenges, consequently requiring new innovations to address (Keles & Keles, 2010). One such innovation, in response to the challenges, is the use of computers (or generally, technology) in education, which has been "*a means of extending sometimes limited reach of humanity*" (Siemens & Tittenberger, 2009, p.12). An early example, traceable to 1959, is Programmed Logic for Automatic Teaching Operations (PLATO)—a computer-based teaching system developed at the University of Illinois (Hickey & Newton, 1967; Suppes, 1979; Woolf, 2009). PLATO constitutes an early intervention in education that exploits several teaching logics, e.g. "tutorial" and "inquiry" logics (Hickey & Newton, 1967). The teaching system contributed to the emergence of computer-based tutoring systems, commonly known as CAI or CAL systems.

CAI—a class of tutoring systems—was frame-based with hard-coded links for instructional purposes (Freedman, 2000; Keles et al., 2009). It presented material in a static and linear manner, in which every student was expected to receive the same courseware. In essence, it was more courseware-sequencing software, but enabled users to have some control over how they navigated system content (Keles & Keles, 2010). Thus, the user's "*...input is not controlled by the computer*" (Carbonell, 1970, p.193). Though CAI was an innovation in education, it had many problems that tended to limit its usefulness (Siddapa, Manjunath & Kurian, 2009). These problems included its static interface, or “storyboard” representation of content (grouped into topics in a curriculum-like format), recipe-like procedures, static/insufficient feedbacks, highly primitive tutoring strategies, non-adaptive tutoring processes, and limited ways of accomplishing a learning task (Carbonell, 1970; Murray, 2003d; Dede, 2008; Siddapa, Manjunath & Kurian, 2009). Also, early CAI assumed the student’s response would always be correct; but when not the case, then the system had to be modified. Later versions of the system provided a form of branching that allowed the system to respond according to the answer provided by students (Carbonell, 1970; Nwana, 1990; Woolf, 2009).

Despite the improvement, CAI feedback on problems had limited learning use. To expand on this, Ohlsson (1996a) argued that the usefulness of the problem-solving approach depended on how much was learned from error. Hence, for feedback to be helpful, a tutoring system should inform a student “*why*” the answer they provided was wrong; it should also state where the error was (Mitrovic, 2012). But in CAI systems, specific feedback to students’ problem-solving strategies and complex pedagogical activities (e.g. diagnosis) were still difficult to achieve. This was because CAI systems lacked the knowledge of the domain in context. Consequently, the system designer needed to handcraft or define all relevant problems, solutions, feedback and all other pedagogical actions (Rickel, 1989; Conati, 2009). In order to address these deficiencies, embedding “intelligence” into CAI systems was conceived. A pioneering work—the initial introduction of Artificial Intelligence (AI) in the classic CAI system—was initiated by, and attributed to Carbonell (1970). This led to the birth of ITS, a class of computer-based tutoring system, but differing from CAI due to embedded “intelligence”.

Although ITS's birth was in the 1970s, it actually gained popularity in the 1990s (Conati, 2009; Siddapa, Manjunath & Kurian, 2009; Li, Zhuying & Bing, 2010) due to the contributions of other researchers. Initially, they defined an ambious goal, mainly the adoption of a human tutor as their educational model, and application of AI techniques to realise this model in “intelligent” computer-based tutoring systems (Corbett, Koedinger & Anderson, 1997; Kay, 1997). The choice of a human tutor, as a model, can be attributed to the realisation that the human one-to-one tutoring strategy provided a highly efficient learning environment for learners (Cohen, Kulik & Kulik, 1982 cited in: Corbett, Koedinger & Anderson, 1997; Nwana, 1990). Also, it has been estimated to increase mean achievement outcomes by as much as two standard deviations (Bloom, 1984).

In recognition of the significance of Bloom's (ibid.) study, several ITS works acknowledged it (e.g. Anderson et al., 1995; Granic, Glavinic & Stankov, 2004; Ainsworth & Fleming, 2006; Koedinger & Corbett, 2006; Mills & Dalgarno, 2007; Lane & Johnson, 2010; Woolf, 2009; Chi & VanLehn, 2010), to the extent that it has been difficult to separate the one-to-one tutoring strategy from ITS construction. Thus, its implementation in ITSs extended the availability of this effective teaching-learning mode. Furthermore, ITS also aimed to "*communicate embedded knowledge effectively, not necessarily in an identical manner as a human teacher*" (Shute & Psotka, 1996, p. 571), although attempting to mimic a human teacher as closely as possible. On that note, subsequent sections review ITS research in the literature, being a sub-theme of the research discussed in this thesis.

2.2 Intelligent Tutoring System (ITS)

ITS—an outgrowth of CAI (Freedman, 2000; Keles et al., 2009)—is also referred to as Intelligent Tutor (Woolf, 2009), Expert System (Span, 1993; Ghaoui & Janvier, 2001), Knowledge-based Tutor or Intelligent Computer Aided Instruction (ICAI) (Anderson, Boyle & Reiser, 1985; Span, 1993). ITS is a computerised learning environment that incorporates computational models derived from cognitive science, learning science, computational linguistics, AI, mathematics and other fields (Graesser, Conley & Olney, 2012), thus indicating the disciplines or perspectives shaping ITS research. As a learning environment, it provides individualised instruction (or feedback) to learners without human intervention, while performing a task (Kay, 1997; Iqbal et al., 1999; Woolf, 2009; Woolf et al., 2009). Self (1999, p.1) defines ITS as *"a computer-based system, which attempts to adapt to the needs of learners."* Also, it is a computer programme capable of instructing a user in an intelligent way (VanLehn, 1988), while Wenger (1987) in Keles et al. (2009) defines ITS as a computer program that uses several technological resources to support the teaching-learning process. According to Freedman (2000), ITS is a broad term, and includes any computer-based program, purpose-built for learning and having built-in intelligence.

ITS offers interactive learning and is assumed to be far superior to classroom-style learning (Murray, 1998) or complements it (to say the least). It possesses the potentials of immersing the student in learning (Woolf, 2009; Amoia, Gardent & Perez-

Beltrachini, 2011). By immersion, it thus presupposes that such systems should possess features that enhance the learner's engagement and motivation (i.e. stimulates learning). These features could include: flagging learning goals and outcomes, providing rewards (intrinsic & extrinsic) and goal-directed feedback, or the educational milieu, among other possibilities (see: Shute, 2008; du Boulay et al., 2010). ITS is considered a shift from traditional instructor-centred to learner-centred tutoring (Murray, 1998; Munoz et al., 2006), putting learners in control of their learning and emphasising learning outcomes, and not the process of education. When combined with the traditional approach to teaching, it could form part of a blended strategy utilisable by teachers to improve the efficiency and effectiveness of educational interventions. Therefore, the above definitions summarises the "what" and "what not" of ITS.

From the foregoing, an application qualifies to be an ITS if it possesses some attributes: 'intelligence' to drive its adaptability to the user's needs, designed for 'learning' purposes, and is computer-based. These three basic-requirements must be satisfied to classify as an intelligent tutor. The role of intelligence in ITS appears to justify why early researchers initially opted for AI methodology as a vehicle to investigate how to actualise the concept (i.e. "intelligence"). The implication is that researchers must think-through the best AI technique (or combination of techniques) that will realise the best learning outcome; and accommodate—as much as possible—other educational factors that matter in a one-to-one pedagogic strategy. This is imperative, since computer-based tutoring systems are meant for educational purposes and the human tutor has been proven to deliver the pedagogic benefit of one-to-one tutoring strategy.

Furthermore, ITS "intelligence" enhances its adaptation capability. Consequently, it is regarded as the only type of tutoring system that "cares" about learners (Self, 1999; du Boulay et al., 2010), because it adapts to their educational needs and promotes "learning by doing" (i.e. enables knowledge construction through active participation in learning activities). Thus, to achieve meaningful and productive intelligence in tutoring systems, educational goals should be taken into cognisance; these include the engagement of learners in sustained reasoning activity, and interaction with learners based on deep understanding of learners' behaviour (Corbett, Koedinger & Anderson, 1997). This raises the issue of how learner behaviour could be modelled in order to satisfy identified educational goals.

Research efforts driven by the above and other unlisted goals, have recorded some notable successes, in tune with the promise of ITS—to make artefacts responsive to human needs and varying conditions, and to revolutionise education (Ohlsson & Mitrovic, 2006; Mitrovic, 2012). These are evidenced as demonstrated by successfully evaluated ITSs that cut across several domains, such as programming, geometry, physics, mathematics (see: Arroyo et al., 2001; Matsuda & VanLehn, op.cit.; VanLehn et al., 2005; Martin & Mitrovic, 2008; Costu, Aydin & Filiz, 2009; Brusilovsky et al., 2010) and also the accounting and finance domain (Kinshuk, Patel & Russell, 2000; Patel, Cook & Spencer, 2003).

Mainly, ITS aims to combine the power of AI, cognitive science, learning science and other related disciplines to provide an effective and intelligently driven learning environment (Graesser, Conley & Olney, 2012). The built-in intelligence drives learning in a way that the system understands what to teach, who to teach, and how to teach (Nwana, 1990). This helps stimulate understanding of the domain being taught and responds specifically to the student's problem-solving strategies (Anderson, Boyle & Reiser, 1985). In order to achieve this, it may employ a range of different technologies through the synergy effect of contributing disciplines, including AI, cognitive science, and education.

In relation to the AI role, ITSs are more narrowly conceived as artificial (or expert) systems designed to simulate intelligent aspects of the human tutor, because they may employ AI technique(s). Use of AI in ITS contributed to what Woolf (2009, p. 4) called an "'inflection point'—*a full-scale change in the way an enterprise operates"*. In the current research context, the "enterprise" refers to the education field—specifically, pedagogy. Apart from AI, he also claimed two other components or drivers contributed to an educational "inflection point" (or educational change). These drivers are cognitive science and the Internet, while education represent the enterprise subjected to change.

On that note, the roles or contributions of all the three drivers to tutoring systems and pedagogy are briefly captured below, where:

- *AI is the science or techniques of building computers to do things that would be considered intelligent if done by people; it helps deepen the understanding of knowledge, especially the representation and reasoning of "how to" knowledge, such as procedural knowledge;*
- *Cognitive science is the research into understanding how people behave intelligently; it helps to deepen the understanding of how people think, solve problems, and learn; and*

- *The Internet, a technology provides access to unlimited sources of information, available anytime, anywhere* (Woolf, 2009, p.6).

The above-mentioned drivers share a powerful synergy; the first two, AI and cognitive science, are regarded as two sides of one coin (Woolf, 2009). Both fields help understand the nature of intelligent action, thus are regarded as collaborative approaches to tutoring systems. While AI techniques are used to build software models of cognitive processes, cognitive science research results are used to develop more AI techniques to emulate human behaviour. AI techniques are utilised in education to model learner knowledge, courseware, and teaching strategies (Woolf, 2009). In addition, Internet technology provides a platform for Internet-oriented applications. It enhances education by closing the gap between traditional educational techniques and future trends in technology-blended education (Brusilovsky, 2001 cited in Tzouveli, Mylonas & Kollias, 2008). Thus, the Internet makes more learning material and reasoning available for longer hours than ever before, and supports more students to learn in less time (Woolf, 2009).

To date, some ITSs driven by the above stated drivers/disciplines have progressed into real classroom use (Graesser, Conley & Olney, 2012), while many still remain in the laboratory where they were created. The impact of the stated drivers and other related disciplines resulted in varied ITS implementation approaches. These include cognitive, constraint-based, simulation-based, game-based, and advisory-based tutors, as well as collaborative systems (Taylor & Siemer, 1996; Martin, 1999; Martin, 2001; Koedindger & Corbett, 2006; Khandaker & Soh, 2010; Johnson, 2010; Eagle & Barnes, 2012; Mitrovic, 2012). In some cases, the Internet driver provides an excellent platform for wider access (e.g. Butz, Hua & Maguire, 2004; Keles et al., 2009). Some of these implementations claimed links to AI and some to cognitive science models, e.g. ACT-R—Adaptive Control of Thought-Rational model (Anderson, 1993a; Anderson, 2007―cited in Graesser, Conley & Olney, 2012); and Ohlsson's performance error theory (Ohlsson, 1996a, 1996b; Ohlsson & Mitrovic, 2006; Mitrovic, 2012). Also, several architectures have been implemented to date (Padayachee, 2002). Therefore, to position current research, the sources of ITS theoretical underpinnings, structures and approaches―so far adopted―are further reviewed below.

2.3 Theoretical Contributions to ITS Research

ITSs are computer-based systems that tend to enhance the teaching-learning process, as indicated by the above definitions. Thus, learning constitutes a key element of the process, and is its essence. Learning is a process for acquiring (or constructing) knowledge and skills, leading to mastery of a domain of interest, and gaining in capacity to transfer knowledge or skill to other areas (Bransford, Brown & Cocking, 2004). Therefore, mastery learning occurs when a learner undertsands prerequisite knowledge/skills before moving to higher-level knowledge/skills (Corbett, 2001). However, how people learn (or master a domain) is influenced by many factors, including the methods, activities, role of the learner (active or passive), reward versus punishment, role of the brain, the learning environment, etc. (Hammond et al., 2001; Bransford, Brown & Cocking, 2004; Wood, 2004; Swan, 2005; Chan et al., 2006). These factors come into play during the learning process, and impact either positively or negatively. As such, they should be addressed adequately in order to achieve the essence of learning.

Conventional pedagogical theories capture these factors, one way or the other, giving meaning or explanation to them, and establishing connections between them. Through theories, one can better understand the implications of the variables involved in a learning process and guide learning activities. Also, theories could assist knowing how to learn in an efficient manner, providing a platform for handling the learning scenario in a way that yields better learning outcomes. Therefore, for better understanding of the role of theory and its boundaries, a working definition would be required to provide clarity and application. This could help shape current research and give clarity to "what is" and "what is not" theory. Although there are several definitions, a good insight into what theory is, was given by Hammond et al. (2001, p.15). They defined theory as:

> *.... a way of thinking and a model of how things work, how principles are related, and what causes things to work together. Learning theories address key questions, for example, how does learning happen? How does motivation occur? What influences students' development? A theory is not just an idea. It's an idea that is a coherent explanation of a set of relationships that has been tested with lots of research. If the idea survives rigorous testing, that theory is said to have empirical grounding.*

The above-stated excerpt provides a working definition of theory, clarifies its boundaries, stating its essence and the how and what of learning. The question now is: from where does ITS derive its theoretical underpinnings? Why is theoretical consideration significant? In order to address these questions, some views were thus reviewed. Hartley (2010) reflected on the relationship between theory and practice, but with respect to ITS's precursor (i.e. CAL—also known as CAI). He noted the influences of theories on CAL. These included the development of systems that generated diagnostic feedback, reinforcement of learning, and the exploration of practical context that enables students to explore, reflect and solve problems. A practical example is CASTE—Course Assembly System and Tutorial Environment—a tutorial system that represents an embodiment of CT in an artefact (Scott, 2007, 2008); this system is further discussed in chapter 3. This system demonstrated the conversation between participants involved in a teaching-learning process. On account of the role of theory in CAL systems (e.g. Pask's CASTE system), Hartley (2010) concluded that there is interplay between theory and practice, especially with respect to educational, computer-based systems. His reflection indicated how theory impacted CAI systems. Therefore, if theory relates to CAI practice, a precursor of ITS, then why not ITS?

The view of Self (1990b, 1994) that ITS theory-practice linkage should be established still appears relevant, considering ITS progress to date. He argued that ITS design does not have any link to a theoretical base (Self, 1990b). As a follow-up, he asked: "*Would* ITSs *ever be built by a blend of beautiful theory and empirical fine tuning?*" (Self, 1999, p.354) Although this view does not deny the existence of theory-practice relationship, as concluded in Hartley (2010), rather, it points to the trend in ITS work. Despite Self's (1990b) stance, he admitted that recent works are beginning to follow the theoretical path. However, he argued that if any theoretical basis is to emerge, it should come from AI, i.e. such formalisation should be psychologically and educationally neutral. This stance seems contestable and could be attributed to the motive of AI researchers in the education field, which is the implementation of AI techniques in educational systems. So, one is not surprised that his view tends towards AI. Nevertheless, his argument in favour of formalisation suggests that most ITS researches might have been undertaken without any formal theory in mind. Rather, they were driven by an informal theoretical foundation or principles (Self, 1990b), significantly drawn from cognitive science. These assumptions or principles, according to Self (1990b), cannot translate into any formal theory. Therefore, steps must be taken to develop ITS theories from AI that will help guide the design/development of ITSs.

Accordingly, the above-stated view appears to be driven by Self's (1990b) AI background/interest. On that basis, one may be content to agree with the argument that ITS has no in-house theory. Thus, theories may be required to drive its design process. However, the argument that ITS theories should emerge from AI can be contentious and tends more towards an AI-driven engineering/design perspective. This seems too narrow to consider for an educational tool, because engineering theories alone, if any exist, cannot produce reliable tutoring systems that are meant for classroom use, or to support learning activities outside the classroom. Moreover, if Hartley (2010) could establish a significant role of pedagogical theories in CAL (a precursor of ITS)—which is also meant for learning purpose, doing same for ITS is favourably arguable. Thus, a broader theoretical source that takes into consideration all relevant disciplines, especially education (the focus of ITS implementation), should be considered for ITS.

Notwithstanding the above arguments, one could still assume that each ITS research was driven by some theoretical assumptions. On that note, one would expect AI and the other two disciplines to contribute theoretically to the field of ITS; or, that the ITS field has matured to have its own in-house theory (or set of theories). Ascertaining the source(s) of ITS theoretical foundation could enhance understanding and value of its researches. The question then is: what is/are ITS theory (or theories)? If none, where does it derive its theoretical foundation from, and what are these theories? Can we integrate theories from all three disciplines to produce reliable tutoring systems that meet educational goals since they are created purposely for education? These are open questions requiring answers in order to determine how ITSs were designed.

In the light of the above, a quick review of the theoretical role of the disciplines shaping the field could be appropriate at this point. This is necessary, in recognition of their contributions to ITS practice, and the need for sound theory-practice relationship and formalisation of ITS design (Self, 1990b, 1999; Hartley, 2010). Thus, a second look into Self's (1990b) argument may be necessary. A case for an educational perspective, rather than an AI-driven engineering perspective, may be given higher priority. An educational perspective coupled with its process could lead to relevant theories that could guide ITS design. This appears viable, since ITS is an interdisciplinary field that could benefit largely from the theoretical foundation of relevant disciplines, especially conventional educational theories. Additionally, many of the pedagogical theories seem to capture variables that play-out in a learning process, which could enrich the design of technology-driven tutoring systems.

Also, a review of relevant disciplines could enhance holistic understanding of their possible contributions. The review could enable the formation of a formidable theoretical foundation that will yield a reliable tutoring system, define a clear direction, and identify the boundaries of each participating discipline. Also, it could enhance the formalisation of the design/development of ITSs. In anticipation of the foregoing, a survey of ITS/Authoring literature was undertaken to ascertain the theoretical contribution of AI, cognitive science and education to ITS research.

2.3.1 Artificial Intelligence (AI) and Theoretical Contributions

Consequent to the above, a survey of the ITS/Authoring literature revealed that the AI field was considered rich in techniques that could be used to implement the concept of "intelligence" in tutoring systems, in an attempt to address problems associated with CAI systems (see section 2.1). Consequently, many research issues emerged, since inception of ITS studies. These include: the learner's plan recognition (Greer & Koehn, 1995; Conati & VanLehn, 1996; Carberry, 2001; Conati, Gertner & VanLehn, 2002; Liu et al., 2011), acquisition of cultural knowledge and communication (Lane & Ogan, 2009), adaptation and personalisation of assessment, control of students' skills and feedback between students and their tutor (Melis, 2005; Vasilyeva, Pechenizkiy & Puuronen, 2006; Vasilyeva et al, 2007; Gladun et al., 2009), mixed-initiative dialogue (Carbonell, 1970; Graesser et al., 2001; Graesser et al., 2005), and decision-making process (Butz, Hua & Maguire, 2004).

Many of the above issues and others unlisted were investigated using AI techniques; for example, BNs—based on probability theory, was utilised as a framework for uncertainty management (see: Conati, Gertner & VanLehn, 2002; Butz, Hua & Maguire, 2004; Schiaffino, Garcia & Amandi, 2008). Also, semantic net (Carbonell, 1970), fuzzy logic (Kharya, Sharma & Thomas, 2010; Zarandi, Khademian & Minaei-Bidgoli, 2012), artificial agents (Lavendelis & Grundspenkis, 2009; Mikic-Fonte et al., 2010; Mikic-Fonte, Burguillo & Nistal, 2012), neural networks (Stathacopoulou, Magoulas, & Grigoriadou, 1999; Baylari & Montazer, 2009), and case-based reasoning (Ciloglugill & Inceoglu, 2010) have all been utilised. In addition, the hierarchical granularity technique has been employed to implement plan recognition (see: Greer & Koehn, 1995; Liu et al., 2011). These techniques were applied in different ways. In some cases, they were applied as a single Knowledge Representation (KR) of various components of ITS, while in other cases, they were used in a hybrid format, i.e. a combination of two or more techniques within an ITS (Stathacopoulou, Magoulas, & Grigoriadou, 1999; Hatzilygeroudis & Prentzas, 2004).

The questions now are: are these techniques theories? Or, what is/are the theoretical foundation(s) of those works that utilised them? Do they really, in a formal sense, have any theoretical undertone? Or, are they underpinned by informal theories? These are open questions. They could provide a window through which the theoretical perspective of AI-based studies can be appreciated; or, to determine whether there has been any theoretical contribution from AI. What could be said is that many of these techniques have some mathematical theoretical undertones, which guide their applicability. For example, BN emerged from probability theory, thus suitable or adaptable to scenarios that match such a mathematical theoretical underpinning. So, one may be tempted to assume that ITS designed with AI techniques are underpinned by some theoretical foundation from mathematics, although indirectly.

Despite their mathematical foundation, these techniques do not identify or capture the variables that should be considered in the teaching-learning process, for learning to be achieved. Thus, they do not appear appropriately placed to inform features that should be considered when designing tutoring systems. However, they may be used to implement ITS features, after they might have been identified by relevant theories. Hence, while AI techniques are desirable in the implementation of ITS design, efforts should be made to underpin ITS design with relevant theories from other disciplines. Such theories should be able to state how learning occurs, the learning context, the variables involved, and how these variables should be treated to achieve effective learning.

2.3.2 Cognitive Science and Theoretical Contributions

Cognitive science emerged as a discipline spanning fields of psychology, AI, philosophy, linguistics, anthropology, and the neurosciences. It holds the premise that cognitive processes are computations. It is a perspective providing the platform for direct comparison of natural intelligence and AI, where emphasis is on a methodology that integrates formal and empirical analyses with computational synthesis (Strube, 2004). To date, cognitive science has practically materialised in the form of computer simulations—a hallmark of the field, and has impacted many ITS works through modelling of cognitive issues (as demonstrated in Aleven & Koedinger, 2002; Aleven et al., 2004; Aleven et al., 2005; Koedinger & Corbett, 2006; Ohlsson & Mitrovic, 2006; Mitrovic & Weerasinghe, 2009; Muldner & Conati, 2010). Many of these implementations were underpinned by cognitive theories (e.g. ACT-R, SOAR, EPIC, CoJACK, etc. – see below for meanings), sometimes referred to as cognitive architectures (Lewis, 2001; Young, 2004; Muller, Heuvelink & Both, 2008; Ritter, 2009; Peebles & Banks, 2010) or principles (Self, 1990b, 1999).

Prominent among the cognitive theories or principles are:

- ACT-R (Adaptive Control of Thoughts—Rational) by John Anderson (Anderson, 1993a; Lovett, 2004; Anderson et al., 2004; Anderson & Lebiere, 1998 cited in Koedinger & Corbett, 2006) – usually implemented as production rules to model domain knowledge and learning behaviour. Note that production rules consititute the expert model of a target domain that can be compared with a learner's solution goal to determine the actualisation of learning. Thus, they are knowledge components that are flexibly recombinable or a way to represent chucks of knowledge and decision capabilities during problem-solving activities (ibid); and
- SOAR (State-of-the Art computational theory) by Allen Newell (Newell & Simon, 1972; Newell, 1994) – a computational theory of human cognition (Lewis, 2001), which defines the world as large problem space with states and goals, and considers behaviour as movement within the space by performing actions, either internal (mental activity) or external (observable movements in the environments) (Muller, Heuvelink & Both, 2008; Ritter, 2009).

Both SOAR and ACT-R model behaviour by reducing it to problem solving, but the former does it explicitly, whereas the latter implies it by being goal-directed (Ritter, 2009).

Also, among the theories are Ohlsson's Performance Error (PE) theory, sometimes referred to as theory of "learning-by-doing", (Ohlsson, 1996a); this is an informal theory, implemented as constraints (or states), usually used to model domain knowledge and student learning behaviour. EPIC (Executive-Process/Interactive Control) by David Kieras and David Meyer (Kieras, Wood & Meyer, 1997 cited in: Anderson et al., 2004;

and Ritter, 2009) is a production system architecture that relates cognition to perception and action (Anderson et al., 2004).

However, among these theories, ACT-R and PE are the most commonly referenced in the ITS literature. Both have been claimed to impact several works (see Koedinger & Corbett, 2006; Ohlsson & Mitrovic, 2006; Mitrovic & Weerasinghe, 2009; Muldner & Conati, 2010). Indeed, they provide a conceptual framework for building models of how people learn in intelligent tutors using production rules and constraints respectively. Both have their advantages, as well as limitations (see Martin, 2001; Kodaganallur, Weitz & Rosenthal, 2005, 2006). Notwithstanding their frequent utilisation in ITS studies, they are regarded as mere principles (Self, 1990b; Dessus, Mandin & Zampa, 2008), and according to Self (1990b), there exists a gap between "theory" and principles and between principles and implementation. Self (1990b, p.3) argued that these "*...principles do not determine an implementation and it is not possible to say categorically whether an ITS has been implemented in accordance with the principles or not.*" Also, Dessus, Mandin & Zampa (2008) noted that these models have drawbacks, including lack of high-level categorisation principles, and their pedagogical or system-relatedness. This could suggest reasons behind the latter's preference for features occurring in the real world of teaching-learning, to develop tutoring principles that could be adopted in ITS.

In the light of the above, formal theories are still required to understand and identify factors that need to be considered when constructing an ITS. The above-mentioned principles (e.g. ACT-R) could be integrated with conventional theories (e.g. Conversation Theory), where appropriate. For example, while traditional theory captures variables (e.g. teaching/learning medium, strategy, styles, etc.) that could inform features that should be incorporated in ITS, an approach based on production rules, as prescribed by ACT-R architecture, could be adopted to represent domain knowledge and to track learning behaviour. In place of the latter approach, the constraints approach promoted by Ohlsson's (1996b) theory could also be implemented. Also, AI techniques could be used to implement some of the principles, where applicable. On that basis, it could be argued that even studies claiming a link to these cognitive theories, still require formal theories in order to formalise the construction of ITS, and to establish the link between theory and practice.

2.3.3 Education and Theoretical Contributions

Education has been the purpose of many ITS implementations. This is evidence when one considers various ITS implementations from their evaluation perspectives, in which educational issues, such as achievement or learning gains, effective learning, etc., were measured (e.g. Kinshuk, Patel & Russell, 2000; VanLehn et al., 2005; Siddapa, Manjunath & Kurian, 2009; Muldner & Conati, 2010; Jeremic, Jovanovic & Gasevic, 2009, 2012). This demonstrates the educational significance of ITS research, moreso that one of its goals is to revolutionise education (Ohlsson & Mitrovic, 2006; Mitrovic, 2012). In that respect, some notable successes had been achieved. Their underlying theoretical assumptions, if any, principles and/or methodologies however, were significantly drawn from disciplines aside from education, e.g. signal processing (Graesser & D'Mello, 2012), cognitive science (Aleven et al., 2009), and AI (Chaouachi et al., 2010; Amoia, Gardent & Perez-Beltrachni, 2011; Chalfoun & Frasson, 2011; Jeremic, Jovanovic & Gasevic, 2012; Piech et al., 2012). One of the reasons that had been proffered is that these works were employed as research platforms to enable rigorous experimentation of the principles or techniques in classrooms, with real students and real courses (Koedinger & Corbett, 2006).

Despite the foregoing, the researchers did not rule out the eventual educational intent, goal or usefulness of their works. Educational perspectives of the latter works appear to give more weight to high consideration of conventional educational theories in the design of ITSs. However, this does not seem the trend as exemplified by many studies (see Lavendelis & Grundspenkis, 2009; Diziol et al., 2010; Gilbert et al., 2011). Moreover, no explicit mention was made in that regard. The studies demonstrated the implementation of various AI techniques and cognitive science architectures/principles, while attempting to investigate various educational-related issues. While these approaches might be desirable, emphasis should be geared towards integrating these techniques/principles with traditional teaching-learning theories. By so doing, various learning factors—captured well by traditional educational theories—could be accommodated in order to design effective tutoring systems.

Considering the latter point, it is worth stating that education has a lifelong history and has developed into a vast and theory-rich field. Numerous ideas have been developed to date, attempting to unravel the concepts of "knowledge" and "truth", and how learning is acquired. These ideas pervade educational works and exist from the time of philosophers to the latest theorists. Some of these ideas have undergone empirical analysis, and confirmed dependable theoretical concepts that could explain pedagogical issues. Due to the theoretical richness of the education field, and the need to develop effective tutoring systems, some researchers have acknowledged the field, to benefit from its wealth of research information and theories. As a result, some elements of these traditional theories were informally referenced in a few ITS studies, such as scaffolding, fading etc. (e.g. Conati & VanLehn, 1996; Siang & Rao, 2003; Lee, Cho & Choi, 2006). Although these studies merely mentioned conventional educational theories, their design was more or less shaped by AI techniques instead of the conventional educational theories. However, they succeeded in highlighting the theory-practice connection, thereby confirming the significant role educational theories could play in tutoring systems.

After due consideration of the various theoretical approaches adopted, this research aligns with the earlier stated argument of Self (1990b, 1999). Also, this work realises the essentiality of the theory-practice interplay suggested by Hartley (2010). Although the former's argument is AI-driven (i.e. psychologically and educationally neutral), Rodrigues, Novais & Santos (2005) argued to the contrary. The latter claimed that psychology and educational sciences are required to develop efficient and effective ITSs. Also, this research argues that an education-driven theoretical framework should be the basis for ITS design because it was considered from educational goal perspectives. Thus, this position enhanced the adoption of theoretical assumptions that emerged from a natural educational setting, as promoted by educational theories. The stance aligns with the argument of Dessus, Mandin & Zampa's (2008), discussed earlier (which favoured the naturalistic approach to cognitive principles-driven design of educational systems).

Thus, the research sees the need to underpin the construction of ITSs using a pedagogic metamodel that emerges from educational theoretical frameworks, Conversation Theory (CT) (Pask, 1976a) and Cognitive Apprenticeship (CA) (Collins, Brown & Newman, 1989). While these frameworks are discussed later in the chapter (see section 2.10 below), the questions that come to mind are: what type of ITSs had been constructed? What type should emerge from the metamodel approach being considered in this research? How could we enhance their production? Answers to these questions require a review of existing approaches to ITS development.

2.4 Intelligent Tutoring Systems (ITS) Approaches

ITSs were implemented using different approaches. Some of the approaches identified in the literature are: cognitive, simulation, coaching, and collaborative approaches; these are further discussed as follows:

- **Cognitive approach** – ITSs based on this approach emphasise learning through cognitive skill development. They present a problem-solving environment with rich feedback (Koedinger & Aleven, 2007). This possesses rich and dynamic models of how students and teachers reason, adapted over time as learners' understanding grows (Corbett & Anderson, 2001; Marlino et al., 2004 cited in Woolf, 2009). The models of student knowledge depict the key ideas learners should understand, as well as common learner conceptions and misconceptions, thus enhancing implementation of a learner-centred approach. Examples of tutors with an underlying cognitive model are LISP tutor and Geometry tutor (Anderson, Boyle & Reiser, 1985; Anderson et al., 1995). The former is used to teach a basic programming construct, assisting students to write short programs in LISP programming language; while the latter assists students to search for geometry proofs and present them in proof-graph form.
- **Collaborative approach**—this class of systems emphasises collaboration between different learners or users of the system, and considers learning as a group process. They facilitate quality interaction between students, encourage participation, support collaborative skills practice and promote group learning. Instead of emphasising direct tutoring of individual knowledge in the domain under consideration, group learning is emphasised (Dillenbourg & Self, 1992 cited in Self, 1994a, 1994b; Dillenbourg, 2002; Pozzebon et al., 2007; Isotani & Mizoguchi, 2008; Diziol et al., 2010). An example of this class of tutor is the collaborative version of Cognitive Tutor Algebra (a cognitive tutoring system for mathematics). Although it was designed as a cognitive tutor, due to its extension with the addition of collaborative features using both "Fixed" and "Adaptive" collaboration scripts in Diziol et al. (2010), it can also be classified as a collaborative system.
- **Coaching approach**—another category of ITS is the computer coach. This tutoring system presents an environment that enables learners to practise varying tasks. It represents an advanced peer that leads learners through deadlocks. Also, coach-based ITSs enable users to overcome problems that would otherwise be

difficult to surmount. Typical examples are: WEST – game-playing coach (Burton & Brown, 1982 cited in: Hsieh & Hsieh, 2001; and Grundspenkis, 2008); SHERLOCK (Lajoie & Lesgold, 1992 cited in: du Boulay & Luckin, 2001; Woolf, 2009); and Andes Self-Explanation Coach (SE-Coach) version (Conati & VanLehn, 2000; Conati, 2009).

- **Simulation approach**—simulation-based ITS presents an environment where students can experiment in the selected domain with guidance from the system (Taylor & Siemer, 1996; Munro et al., 1997; Johnson, 2010). Examples are: Cardiac Tutor – provides intelligent simulation to help medical personnel learn procedures (Elliot, 1996 cited in Woolf, 2009) and SHERLOCK tutor (mentioned above—also implemented this approach).

Although ITS can be classified according to the approach implemented, in reality, it is not uncommon to have ITS with varying features falling into various classifications. An example is SHERLOCK—a tutor that implements both simulation and coaching approaches. Thus, the major issue is to determine the driving goal for the design and development of an ITS, because it influences the features composition of the tutoring system, thereby determining the approach to be implemented.

Therefore, in this research, focus is on ITSs that adopt a cognitive approach, and implement a problem-solving approach that enables the development of procedural skill, which is required to gain mastery of numerical aspects of applied numerical disciplines. The implementation should be undertaken within a non-verbal conversational learning environment that supports cognitive activities and enables the construction of knowledge. The foregoing was considered because: [i] the research is contextualised within the applied numerical domains and requires procedural knowledge/skills through problem solving to master; [ii] its problem solving, conversation and cognitive learning strategies align with the philosophical assumptions of the theoretical frameworks that underpin this research; and [iii] the accounting domain, which is the implementation domain for this research, involves categorisation and application of rules.

So, the current research aims to test the cognitive-coaching approach through the conception and implementation of a pedagogic metamodel (as mentioned earlier), which will be used to generate intelligent tutors in the applied numerical domains. Such intelligent tutors should present problems in the target domain, thereafter provide support and appropriate feedback as the student engages with it during the learning process. However, this would only be possible if the intelligent tutor is composed of the appropriate modules, because ITS uses knowledge captured in its components to drive learning intelligently (Shute & Psotka, 1996; Keles & Keles, 2010). Thus, a review of these components is considered essential in order to determine the ITS structure that should be adopted in this research.

2.5 Components of Intelligent Tutoring System (ITS)

An ITS is only as effective as the various components that constitute it, i.e. the degree of functionality and effectiveness in enhancing learning varies according to the relative level of intelligence built into the various constituent components. This relativity is captured by Freedman (2000, p.1):

> *... a project focusing on intelligence in the domain model may generate solutions to complex and novel problems so that students can always have new problems to practice on, but it might only have simple methods for teaching those problems, while a system that concentrates on multiple or novel ways to teach a particular topic might find a less sophisticated representation of that content sufficient. When multiple components contain intelligence, homogeneous or heterogeneous representations can be used.*

Hence, building an ITS needs careful preparation in terms of describing the knowledge and possible behaviours of its underlying components. This description needs to be done in a formal language in order that the ITS may process information and draw inferences aimed at generating feedback or instruction. However, a mere description is not enough; the knowledge contained in each component of the model should be organised and linked to an inference engine. It is through the latter's interaction with the descriptive data that tutorial feedback can be generated.

In recognition of the above, several architectures have been proposed in the literature, with a varying number of components or modules, namely three, four or more components (Self, 1999; Padayachee, 2002; Keles et al., 2009; Siddapa, Manjunath & Kurian, 2009; Li, Zhuying & Bing, 2010). For example, though Woolf (2009) aligns with a four-module structure, the latter argues that some combination of these modules exists due to their overlapping functionalities. The differences in structure may be attributed to functionalities, and breadth and depth of each ITS concerned. For now, there is no agreed standard framework that stipulates what should constitute an ITS; each project determines what best fits its goal(s). However, some basic functionalities have been identified in the literature. As a result, a classic ITS is generally conceived as consisting of four main components, namely: domain module, student module, teaching/pedagogy module, and a learning environment or user interface (Freedman, 2000; Hsieh & Hsieh, 2001; Yang, Kinshuk & Patel, 2002; Padayachee, 2002; Samuelis, 2007; Woolf, 2009; Li, Zhuying & Bing, 2010).

Woolf (2009) distinguishes between the terms "module" and "model" and chooses to use a broader term "communication module" for "learning environment" or "user interface". A component (or module) is conceptually different from a model, although the former (i.e. module) and the latter (i.e. model) refer to a similar object. Woolf (2009, p.49) distinguished between both terms, where:

> *"....a module of a tutor is a component of code that holds knowledge about the domain, student, teaching and communication"; while "a model refers to a representation of knowledge, that is, the data structure of that module corresponding to the representation used to summarise the data for purposes of description or prediction. "*

Thus, the words "component" and "module" are interchangeably used to refer to the same concept in this discussion, but different from a model. Also, the word "communication" is used interchangeably with "user interface", although communication may have a broader meaning and implementation depending on each ITS design.

In the light of the above, one could conclude that teaching and learning activities can be adequately captured with the domain, student and teaching modules, since these modules touch on the main ingredients of pedagogy – what to teach, who to teach, and how to teach respectively. Also, there should be a medium of communication between the learner and the domain expert. In that regard, this research adopts the conventional four-module structure, which is discussed in the following sections.

2.5.1 The Domain component

The domain component, also termed an expert system, expert module, or domain module, constitutes the facts and rules of the chosen area of study to be taught to the learner. It includes the knowledge concepts, referred to as the basic entities that constitute the target domain. Thus, it represents the knowledge of the domain expert, a person expected to have acquired long years of experience in the domain and having in-depth knowledge of possible grey areas and ways of tackling them. In essence, it answers the question of "*what to teach*", the source of knowledge for other components of ITS, thus it is regularly invoked (Li, Zhuying & Bing, 2010). Building a domain module is considered to be the first step towards representing student knowledge, although the latter might represent the same knowledge as the domain and solve the same problems (Woolf, 2009)—since both modules overlap in functionalities. In order to build a domain module (or expert system), a number of methods could be used to simulate the performance of the human expert. Common to most or all are:

i. the creation of a knowledge base using the KR formalism to capture the domain expert's knowledge; and
ii. using a knowledge-gathering process to gather domain expert knowledge and codify it according to the formalism, referred to as knowledge engineering.

In the literature, varied domain KR formalisms had been used, classified as either single or hybrid KR. Hybrid formalisms are KR schemes that integrate two or more single KR formalisms (Hatzilygeroudis & Prentzas, 2004). In the class of single representation formalism utilised in ITSs are: semantic networks, frames (or schemata), symbolic rules, fuzzy logic, BNs, and case-based reasoning among others (Carbonell, 1970; Baylari & Montazer, 2009; Kharya, Sharma & Thomas, 2010; Ciloglugill & Inceoglu, 2010; Zarandi, Khademian & Minaei-Bidgoli, 2012). While hybrid schemes utilised include: neuro-symbolic and neuro-fuzzy KR. These formalisms are drawn from the field of AI. Other forms of KRs exist, derived from the field of cognitive science, namely ACT-R based production-rules (Taatgen, Lebiere & Anderson, 2006; Taatgen & Anderson, 2008, 2010) and Constraint-Based Modelling (Ohlssons, 1992; Ohlsson & Mitrovic, 2006; Ohlsson et al., 2007; Mitrovic, 2012).

Within ITSs, the domain component plays some important roles. One, it serves as the source of knowledge to be presented to the learner, which includes generating questions, explanations and feedback. Two, it provides a standard for evaluating learner performance and updating the student model with the learner's behaviour/performance. In order to accomplish the evaluation task, it should generate comparable solutions to problems in the same context as the student, and detect common systematic mistakes and any gap in the learner's knowledge structure that may be responsible for it. Also, to supervise the learner's problem solving skills effectively, it should be able to generate sensible and possibly multiple solution paths so that intermediate steps can be compared during the learning process. Equally, the assessment of the learner's overall progress can be achieved via this module, but will require the establishment of some criteria that form the basis for comparing knowledge (Nwana, 1990; Li, Zhuying & Bing, 2010).

Central to ITS in general and domain and/or student components specifically is the domain model. Although, domain model and domain module may be referring to the same object, as stated earlier, they are conceptually different. Consider a domain module as a “wrapper” for domain knowledge codes, while the domain model represents knowledge structure (or formalism or codes or models) of the study area (or domain), which an intelligent tutor is meant to teach (such as accounting, engineering, mathematics, etc.). Its implementation, whether it is situated in the domain module or student model within an ITS design, may take different forms depending on a number of factors. These may include the ITS approach adopted, KR formalism utilised, the subject domain being represented, whether the domain is structured or ill-structured, and the level of granularity of the knowledge to be represented. For example, in cognitive tutors – where emphasis is placed on production rules for guiding problem solving tasks—the domain model may be implemented as a model of low-level production rules (Peebles & Banks, 2010), while constraint-based tutors represent the domain model as a collection of constraints (Martin & Mitrovic, 2000; Menzel, 2006; Martin, Mitrovic & Suraweera, 2008).

During tutoring when a call is made for domain knowledge, the domain module references the domain model (or expert system/model) for detailed information of knowledge or behaviour of the chosen domain to be taught. It derives information from the domain model to provide guidance on problem selection and generation, responds to the learner with adequate feedback, clarify uncertainties, which may require one or more human experts’ consultation, and updates the student model with the learner’s behaviour. So, in this research, the domain module assumes the meaning and implementation discussed above. However, its knowledge is captured in the form of a network of interconnected variables, in which each variable is derivable from one or more rules. This is informed by one of the theoretical frameworks (i.e. CT — Pask, 1976a) that underpin this research, which will be discussed later in the chapter.

2.5.2 The Student component

The student module—also known as the learner module—is an essential unit within an ITS. Construction of any ITS revolves around it, because it is learner focused. A learner needs an adaptive and personalised system with an effective feedback/help mechanism to achieve his/her learning goals. A system that can provide such features must dynamically construct the learner's model during the tutorial session, a requirement for driving the system interactivity and response. This model of the learner, known as the student model, is contained within the student module. So, the student module utilises the student model to communicate with other modules in an intelligent tutor. The student model normally contains the description of student knowledge or behaviours, including learner misconceptions and knowledge gaps. In summary, it is regarded as an up-to-date knowledge state of a student. Thus, the student model addresses the question of "who is learning" (Li, Zhuying & Bing, 2010). So, in tutoring systems that do not have a domain module, the student module may also house—within it—the domain model. This is so, because the student model is dynamically created and updated (Brusilovsky, 1994), and cannot function alone, because the student module requires information from two sources: the learner's interaction at runtime, and the stored domain model, in order to create or update the student model. This information put together can then be used to construct the student model at the start of learning, and update it as learning progresses.

Like the domain model, the student model is conceptually different from the student module, although both refer to the same object—the learner. The student module generates learner models that are used as patterns by other system modules. On the other hand, the student model functionally examines the student's reasoning during the learning process, identifies the exact point at which the student went astray, diagnoses the reasons for the error or misconception, and suggests viable ways of overcoming the impasse (Woolf, 2009). In order to achieve these functions, the system should monitor the student's behaviour, and capture a representation of his/her cognitive and affective knowledge in a process called student modelling, in a realisation that affect is intertwined with cognition to guide rational behaviour (Woolf et al., 2009). However not all ITSs capture all aspects of a learner; some are concerned only with the cognitive aspect of a learner, while some attempt to diagnose the effective and motivation aspects of learners (du Boulay et al., 2010). Nevertheless, the process of monitoring student behaviour is common to all, and vital to the functioning of an ITS, since the essential goal of AI in education is to support students with varying learning abilities, disabilities, interests, backgrounds and other learning issues (Shute, 2008). This cannot be achieved unless an ITS supports features that enable modelling of the learner.

Many works emphasised student modelling (as in VanLehn, 1988; Self, 1990a; Baker, Corbett & Koedinger, 2004; Pena, Sossa & Gutierrez, 2007; Perez-Marin & Pascual-Nieto, 2010), such that one wonders why such efforts were channelled towards this aspect of an intelligent tutor. An overview of the student model thus indicates its centrality to the effective functioning, as well as a major model that strengthens the intelligence of a tutoring system. Self (1994a, 1994b) provided the rationale for the inclusion of student modelling in a tutoring system, an attempt to provide a clue to its importance. The latter argued that in the absence of a student model, a computer-based tutoring system will perform in exactly the same way for all users, since there is no basis to behave otherwise. However, we must recognise the fact that students are different, and do have varying preferences (Alves, Pires & Amaral, 2009; Alseddiqi & Mishra, 2010; Hou et al., 2010). As such, they do have: different prior domain knowledge, different interests, different learning aptitudes, etc. So, the aspect of an ITS that takes on board all these different user qualities is the student model; hence, the need for its implementation.

Furthermore on student model functions, Self's (1988) paper, entitled: “Student models: what use are they?”, identified twenty functions (cited in: Nwana, 1990; Patel & Kinshuk, 1997). These functions, as mentioned in Riccucci (2008), can be grouped into six categories:

- Corrective – to repair student misconception or misunderstanding by way of identifying the gap between the student’s knowledge and the domain/correct knowledge, then informs the other sections of the system.
- Elaborative – to help correct “incomplete” student knowledge through knowledge extension; this can be achieved when the model identifies areas where the student requires new material, or a refinement of his/her current understanding of the subject.
- Strategic – to help initiate change of tutoring strategy, rather than using corrective and elaborative approaches—as mentioned above; in this case, the student model will have to provide more information about the learner with respect to current tutoring strategy, as opposed to previous strategy.
- Diagnostic – analysis of the student’s knowledge state in order to identify bugs in knowledge. In some sense, this implies all aspects of student modelling can be subjected to a diagnostic process. In that case, the student model can be used to

refine information about the learner, in order to arrive at a decision. For example, if the ITS wishes to introduce a new topic, but the student model is unable to indicate whether the current level of understanding of the student is adequate, the student model can be requested to generate diagnostic samples for the student.

- Predictive – to help anticipate the student's likely response to the tutorial action. In this case, a student model can act as a "simulator," aimed at simulating the student's behaviour.
- Evaluative – to enable the assessment of student's achievement level. In order to achieve this, the system may have to aggregate across information in its possession.

So far, in implementing student model functions, three types of approaches have been used. Each approach established a link between the student model and domain model. In order to illustrate this, student models are dynamically constructed and updated in relation to the domain being taught, as the student progresses in learning. The constructed student models ultimately depend on the modelling approach implemented in the intelligent tutor. The resulting student KRs or models can then be used to tune system behaviour in relation to the domain. Below, the three implementation models identified in the literature are reviewed.

- **Overlay model** (Brusilovskiy, 1994; Beck, Stern & Haugsjaa, 1996; Smith, 1998; Woolf, 2009; Li, Zhuying & Bing, 2010; Shute & Zapata-Rivera, 2012) – this uses techniques for describing a student's problem-solving skills in terms of a programme designed to be an expert for the chosen domain (Carr & Goldstein, 1977). It assumes that the student's knowledge is a subset of the expert knowledge (see figure 2.1 below), and the goal of tutoring is to enlarge this subset (Smith, 1998). The student model is an overlay on the expert program, in the sense that differences between the student's behaviour and the behaviour of the expert model can be explained by the lack of skills on the part of the student. The approach has some shortcomings: students often have knowledge not included in the expert knowledge; misconceptions are not catered for; the lack of alternative representations for students growing in knowledge or mental models; and there is no way to distinguish between knowledge the student has not grasped, and that, which he has not been exposed to. This is particularly addressed by the differential model (Smith, 1998; Woolf, 2009).

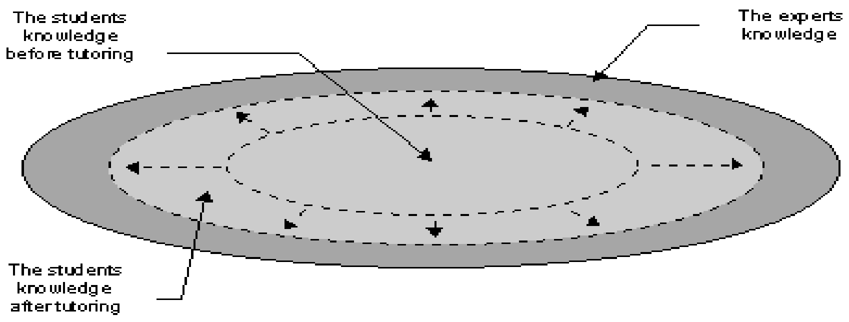


Figure 2.1: An overlay student model
Source: Smith (1998)

- **Differential model** (Brusilovskiy, 1994; Smith, 1998) – this is regarded as an extension of the overlay model in that knowledge is divided into that which the student is exposed to, and that to which the student has not (see figure 2.2 below). While the overlay model is only about the knowledge presented to the student, the differential model includes knowledge not presented; hence, for this reason, it is an extension of overlay model. Like the overlay model, it has the shortcoming of not catering for misconceptions or bugs (Smith, 1998).

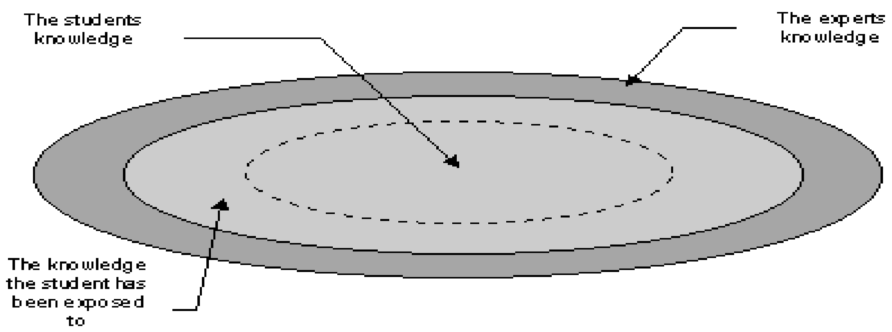


Figure 2.2: A differential student model
Source: Smith (1998)

- **Perturbation models** (Brusilovskiy, 1994; Smith, 1998; Grundspenkis & Strautmane, 2009; Woolf, 2009; Li, Zhuying & Bing, 2010) – also referred to as buggy models, cater for student knowledge that is not part of the expert model (see figure 2.3 below). The difference is on the basis of small perturbations between the student knowledge and some of the expert model's knowledge. An example is lack of knowledge, such as common bugs or misconceptions. Thus, the model extends the expert knowledge with bugs (the difference between expert knowledge and student knowledge). Similar to the overlay model, the goal is to grow the student's subset of expert knowledge, while eliminating bugs (see Beck, Stern & Haugsjaa, 1996).

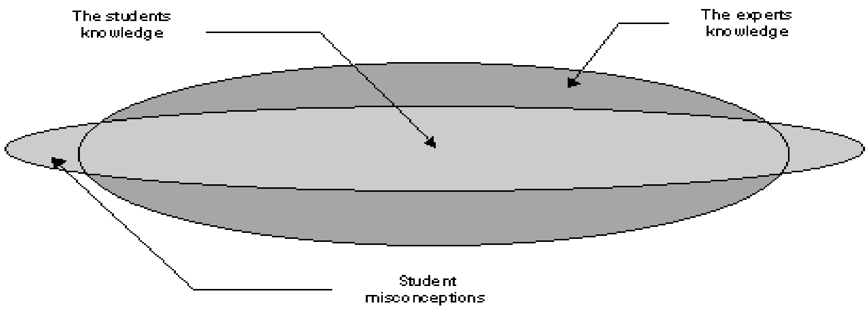


Figure 2.3: A perturbation student model
Source: adopted from Smith (1998)

Similar to the domain model, a number of KR techniques have been employed in order to implement the above approaches in ITS student models (Dillenbourg & Self, 1992, cited in Self, 1994a, 1994b). Some of these techniques were drawn from AI, as well as cognitive science. Like the domain model, student models benefited equally from the implementation of the single knowledge formalism or technique. Also feasible is the implementation of hybrid AI and cognitive science techniques in student models (*e.g. add a Bayesian belief network to a model-tracing tutor*—(Woolf, 2009, p. 80).

Stellan Ohlsson named student modelling "cognitive diagnosis", since the essence of the student model is to know something about the cognitive state of learners—what they know, how they think, and preferably how they learn (Ohlsson, 1986 cited in: Mark & Greer, 1993; Kinshuk, 1996; Dessus, Mandin & Zampa, 2008). Hence, a student model from a "cognitive diagnosis" view-point should include: performance measures, which indicate the proportion of the subject matter known by the learner,; error descriptions, which represent the distorted or misconceived "knowledge units", and simulations, which are executable and enable predictions to be made about learner performance.

One critical issue relating to student modelling that needs consideration is whether the learner really knows a skill after only demonstrating it in an isolated context. The opposite question can also be an issue for consideration, or can it always be true that some piece of knowledge is absent, just because it has not been utilised in certain circumstances. One can assume that incorrect or suboptimal behaviour may be due to incorrect versions of the target knowledge rather than from incomplete knowledge (Kinshuk, 1996). As an example, Brown & Burton (1978) implemented the buggy or perturbation approach utilising a formative student model, an explicit representation of the learner's incorrect versions of the target knowledge for remedial purposes. Consequently, the model must be able to indicate the abilities of learners in relation to the domain being taught and student preferences for any specific tutoring method.

However, this research adopts the overlay approach despite its shortcomings, in realisation that none of the approaches is without a weakness, and may not be feasible to have a system that perfectly matches a human tutor. However, in line with the formal (theory based) approach adopted, it assumes that the domain expert knowledge encompasses the problem posed, and should be sufficient to cater for any misconception within the domain boundary. Thus, to stabilise a tutoring system, problem templates should be structured in such a way that captures expert knowledge and should sufficiently encompass the required problem-solving knowledge. Also, the KR scheme for the student model should be informed by the theoretical framework that underpins this research, in the same way as the domain component.

2.5.3 The Teaching component

The teaching component, also called the tutoring module, instructional module, or pedagogical module in some studies (Padayachee, 2002; Rodrigues, Novais & Santos, 2005; Conati, 2009; Li, Zhuying & Bing, 2010), represents the applicable teaching strategies of an ITS. It models the teaching styles that can be applied to the student depending on the learning context. It is closely connected to the domain and student modules, since these modules can achieve little on their own. The ITS depends on the teaching module to describe how to represent and reason about the domain and the student's interaction with the system. A mismatch between a student's behaviour or knowledge and the domain expert's presumed behaviour or knowledge is signalled to the teaching *module*, which subsequently takes corrective action, such as providing feedback or remedial instruction (Li, Zhuying & Bing, 2010). In order to achieve this, it needs information about how a human tutor would resolve such a conflict, represented in this case by the teaching module.

A variety of teaching approaches have been implemented in the teaching component of ITSs, depending on the context of implementation. One form of implementation is where the ITS monitors the minute activities of learners, adapts system responses to learners' interaction, and still does not release control. So, the learner does not dictate how the system reacts to his activities. Examples of this type of implementation are: QUEST (Frederiksen, 1988 cited in Hsieh & Hsieh, 2001)—a tutor that teaches how to troubleshoot an electrical system; and STEAMER (Holland et al., 1984 cited in: Hsieh & Hsieh, 2001; Woolf, 2009), which teaches how to operate the steam plant in a ship.

On the other hand, some ITSs implement the discovery-learning strategy, but provide guidance during the learning process. In this case, learners have full control of their learning process, but the system can impact the course of action by modifying the learning environment or interface if necessary; included in this category are: PROUST (Soloway & Johnson, 1984 cited in Le & Menzel, 2008) – a system for diagnosing non-syntactic student errors in Pascal programs; and WEST (Brown & Burton, 1982 cited in Grundspenkis, 2008), which renders online coaching for a mathematics game. There is a category of tutors that share control between the system and the learner while in a dialogue. The system adapts based on question and answer exchanges. Examples of such a system are: WHY (Collins & Stevens, 1982 cited in: Nwana, 1990), which teaches the principles of rainfall and corrects learner misconceptions (Hsieh & Hsieh, 2001); and SCHOLAR (Carbonell, 1970) – a geography tutoring system.

While the above discussion is based on the level of control between the system and the learner, another form of implementation based on information availability has been found in some other ITSs. One such implementation is systems that emphasise content presentation. In this category lies the adaptive hypertext systems. Examples are: ISIS-Tutor (Brusilovskiy & Pesin, 1994, cited in Brusilovskiy, 2003) – an intelligent hypertext learning environment designed for learning the print formatting language of an information retrieval system; and ITEM/PG (Brusilovsky, 1993 cited in Brusilovskiy, 2003) – an intelligent tutor environment and manual for introductory programming.

Another category contains the cognitive tutors developed by researchers at Carnegie Mellon University. The ITSs are based on the psychological model of cognition, the *ACT-R – adaptive control of thought rational*, developed by Anderson (1990, 2007)**—**cited in Graesser, Conley & Olney (2012), as well as in Olney, Graesser & Person (2012). Their teaching component implements the concept of model-tracing (i.e. tracing and comparing student knowledge with expert knowledge), to determine the problem difficulty, depth of feedback to give etc. Examples are: PAT (Pump Algebra Tutor), Geometry tutor (Anderson et al., 1995; Koedinger et al., 1997; Ritter et al., 2007); and Andes, Atlas, and Why2-Atlas – physics tutors (Schilze et al., 2000; VanLehn et al., 2002; VanLehn et al., 2005; VanLehn et al., 2007; Woolf, 2009; Mitrovic, 2012). Similar to cognitive tutors are a group of tutors, based on CBM first proposed by Ohlsson (1992) and later extended by Ohlsson & Mitrovic (2007). This group of ITSs model knowledge is in the form of constraints which are utilised to drive teaching, unlike cognitive tutors that were driven using production rules.

In line with earlier argument for theory-based formalisation, such design approach should dictate the teaching strategies that should be implemented, since pedagogy theories do have provision for teaching-learning strategies (e.g. conversation and “teachback” strategies of CT – Pask, 1976a). Thus, this research considers teaching strategy that is similar in functionality to model-tracing (mentioned above). However, the research goes a step further, by considering detail and step-wise tracing of the problem solving process, in order to enhance cognitive visibility, aligning with its underpinning theoretical frameworks (which are discussed later). Step-wise knowledge tracing is known as 'process monitoring' in this research. It enables the investigation of cognitive visibility in relation to the detection of learner misconception and feedback generation that are essential to the success of any ITS that provides a problem-solving learning environment (see Melis, 2005; Shute, 2008). It also means that learning effectiveness can be evaluated.

2.5.4 The Interface component

The interface component, also known as the learning environment or communication module (Padayachee, 2002; Samuelis, 2007; Woolf, 2009; Li, Zhuying & Bing, 2010), is the view through which a student establishes communication with other components of the tutoring system. It establishes a bi-directional communication between the learner and other components within the ITS, and translates between the system’s internal representation and an interface language that is understandable to the learner.

According to Woolf (2009), even with a sound student model and teaching knowledge, an ITS is of limited value without effective communication strategies in place. A confusing or difficult interface, or unattractive feedback platform, will render the ITS ineffective for tutoring purposes. A computer interface has a crucial impact on learning outcome, and for many users, the interface is critical to their interaction, not the computational activities performed beneath the surface (Twidale, 1993 in Woolf, 2009). Due to the importance of this component within an ITS, it may make or mar the tutoring system. Hence, significant effort should be spent in interface design and development, and attention should be paid to qualities, such as ease of use and attractiveness, for learners' acceptance of the tutoring system.

Human tutors use communication to impact knowledge, motivate and engage students during the learning process, and understand students' knowledge. Also, when students develop good communication skills, it impacts positively on their engagement in group learning, knowledge sharing with colleagues, critical thinking, and self-explanation skills (Woolf, 2009). In order to actualise sound communication in a tutoring system, a mirror of good human tutor communication that enhances teaching and learning becomes essential.

Table 2. 1: Human Communicative Strategies Implemented in Intelligent Tutors

Human Communicative Strategies	**Strategies Implemented in Computer Tutors**
Compose explanations spoken or textual; deliver critiques and maintain a mixed initiative dialogue	Atlas, Geometry Cognitive Tutor, AutoTutor
Analyze a student explanation, spoken or textual; question student's approach	Automatic essay analysis/grading (AutoTutor), Geometry Cognitive Tutor
Interpret student formulas or graphics	Free-body diagram (Atlas); interpret formulas (Atlas)
Recognize student's affect (emotion, focus of attention, or motivation)	Interpret speech and visual cues; gesture analysis, face detection; recognise frustration.
Engage students in role play; hire partners for training interactive skills	Virtual humans (Steve); animated pedagogical agents (Herman, Cosmos), interactive simulation (REA)

(**Source:** Woolf, 2009, p. 138)

Human tutors select strategies based on several factors to communicate effectively. They use methods such as: analysing written work, providing explanations/critiques, drawing graphics etc., to communicate. They are able to identify students engaged in learning—through their actions, such as note taking, questioning, contributions etc.—and those not ready to learn. Many strategies used by human tutors have been implemented in intelligent tutors. Some were derived from observation of human tutors and others from technology-based opportunities (virtual learning environments, animated pedagogical agents, etc.), which are unrelated to classroom observation (Woolf, 2009). Table 2.1 above captures the strategies that have been utilised in ITSs. However, the interface component is not the main focus of this research effort, although it played a significant role and attention is accorded it in many previous works, and in this research too. Despite that, efforts are made to take on board best practices—related to this module—in the implementation phase of this research. These efforts resulted in the investigation of usability issues in this research work.

2.6 Intelligent Tutor Systems (ITSs) in the Numeracy Domain

Below some existing ITSs are discussed. They were selected to reflect how the field has progressed towards the main goal of modelling "intelligent" behaviour in educational tutoring systems. Also, because these ITSs were implemented within numerical subjects, and some have been empirically evaluated and have gone beyond the laboratory into real educational use, we hope their inclusion and discussion will help advance work in the numerical domain, which this research aims to accomplish.

2.6.1 Geometry Tutor

A tutoring system in the numerical domain of mathematics was initially developed by John Anderson, but later, advanced versions were produced (Anderson, Boyle & Yost, op. cit; Zelhart & Wallingford, 1994; Matuda & VanLehn, 2004, 2005). It implements a tutorial approach, which supports *'learning by doing'* and visualisation of learning activity as the student makes progress. The system was designed to provide support in performing geometry proofs. It is premised on two assumptions derived from close observation conducted on a human tutor, that the student learns domain-specific problem solving skills by practising the skill; and that the student learns problem solving skills much more effectively if he or she gets support from a full-time human tutor, who is an expert in the domain, than if they learn through the traditional classroom approach.

The ITS is a tutorial system meant for the classroom setting, underpinned by the assumption that the student would have gained basic conceptual knowledge of the domain from a human tutor. So, the system is merely used for tutorial purposes only, enhancing acquisition of practical proof skills. Hence, the system did not capture the domain knowledge of mathematics. Instead, it emphasises the capturing of tutoring rules, which are represented as models. These models are based on underlying assumptions that qualify an *'effective'* tutorial tutor. One, that the tutor must have an internal model of how the skill should be performed, and must be accessible as and when required—represented as the "*ideal*" model. Two, it must anticipate the types of errors a student might commit in the process of learning. These errors are captured in the form of the "*buggy*" model, as proposed by Brown & Burton (1978), to supplement the internal (i.e. ideal) model (Anderson, Boyle, Yost, 1985). These two models are instantiated as production systems derived from the psychological analysis of skills based on the ACT-R production system (Anderson, 1963).

One strong feature that the system provides is visualisation capability. The ITS visualises each problem posed to the student in a diagrammatic form. The student proof steps are equally visualised and updated regularly, thereby enabling a visual feel of steps taken. This provides better comprehension of the problem, and the implications of the steps taken—so the student can determine if he/she is on the right track. Equally, visualisation eliminates the possibility of working on a problem incorrectly, as a result of a wrong drawing, since the drawing aspect is taken over by the system, in essence, limiting what the student can do, to only cogent activities that will enhance learning of the geometry proof.

The system allows the student to work bi-directionally, i.e. the student can work from a goal to known elements, and from known elements to the goal, or a mix of both directions as the problem dictates and the student may wish. This approach alleviates student fears about proofs, and support flexibility. It provides a free learning environment, where the student determines how to approach a problem without being restricted, thereby emphasising skill acquisition rather than how the proof is accomplished. The system also supports some command functions, which form part of the proof language of geometry and are accessible via menu options. When an operand is selected, the system performs a cross-check to confirm validity, attempts execution of the function, and allows the student to respond with input, which is equally verified. Thereafter, an update of both the problem and proof display is carried out.

Student proof moves and student model combined, dictate how the feedback mechanism is instantiated and the type of feedback that is generated. For example, if a student move is in a direction that will not yield the target goal, the system initiates its feedback mechanism—rewarding as well as redirecting the student. The system rewards the student by acknowledging and reassuring them on the effort taken. It then redirects the student on a new path that can lead to the problem goal, or reframes the feedback in the form of a reminder, if the concept is understood, but merely misapplied. Aside from the feedback facility, the system provides a help facility, such as a 'context sensitive' help feature. The student highlights an unknown element of the problem, then the system conducts a search based on it; the system then returns information on the searched element. This is a good facility, which enhances learning and eliminates time that may be wasted if the students were to do the search manually. Help availability and ease of use enhances engagement with the system, even though the student may not understand or remember all the concepts required to carry out the geometry proof; the belief that help is available encourages the student to carry on (Zelhart & Wallingford, 1994).

Despite the huge success of this system, some issues still need to be addressed to advance its usefulness for learning. According to Zelhart & Wallingford (ibid), the system assumes that students have conceptual knowledge of the domain before engaging in practice with the system; so, it is not meant to teach a concept per se, but is a tool for tutorial practice. During the tutorial session, the system cannot account for the time lapse on the part of student. So, if the student is stuck, there is no way the system can determine the current state of the student; this area needs to be addressed.

Equally, students do not like to seek help, neither do students want to be told that they are wrong; hence, the system must incorporate a monitoring scheme that allows it to predict and anticipate any problem a student may have in advance. Since the system allows the student to commit errors and generates feedback based on the error and the student's model, it cannot anticipate when the student is contemplating on a problem, or when the student has a clue, and when the student is afraid to request help. Therefore, improvement of this system must take these factors into consideration, because failure can cause the system to enter into an endless loop (ibid). It also lacks pedagogically-driven theory-based formalisation; instead, the ITS is based on some intuitive assumptions derived from observations of a human tutor, as mentioned above. It does confirm also, the need for formalisation using education theories. Such a theoretical approach has the potential of shaping the design of a good educational system (VanLehn, Jones & Chi, 1992).

2.6.2 Andes Tutor

Andes is an intelligent tutoring system developed to help solve physics homework problems (Gertner & VanLehn, 2000; Graesser et al., 2001). Andes's design is fundamentally underpinned by four principles: to encourage the student to construct new knowledge, which is facilitated through provision of hints that require the users to derive most of the solutions by themselves; to enable knowledge/skill transfer from the system by making its interface much like a piece of paper; to offer immediate feedback after each action, in order to maximise learning opportunities and minimise the amount of time spent on wrong paths; and to offer flexible problem solving order so that solution steps can be performed in the order the student wishes, with the flexibility to skip step(s) as appropriate (Gertner & VanLehn, 2000).

In line with its design principles, it aims to: replace the pencil and paper mode that students employ, while practicing problem solving in physics; allow students to draw diagram(s), unlike Geometry Tutor—where the diagram is automatically drawn by the tutoring system (Anderson, Boyle & Yost, 1985; Zelhart & Wallingford, 1994); define variables and enter equations with the same freedom exercised when using pencil and paper; offer visual feedback—not available in pencil and paper mode—for both correct and incorrect inputs by turning green or red respectively. This type of feedback is classified as the immediate feedback approach, and has been found to enhance learning from other similar studies (Anderson et al., 1995). Also, interaction between the tutoring system and student is driven by the coached problem solving technique—a cognitive skills-teaching technique involving collaboration between system and student to solve physics problems. However, the degree of collaboration between the system and student changes according to the student's progress.

It should be noted that the system also accommodates problems that only need a qualitative solution; this implies that a student would not be required to write any equation. Qualitative reasoning has been found to be useful in deepening understanding; it uncovers students' misconceptions when combined with quantitative reasoning, more than when only the latter is used (Gertner & VanLehn, 2000). On the other hand, providing a quantitative solution within Andes requires a series of actions involving drawing, variable definitions and entering of equations. This drawing capability and flexibility uniquely distinguishes the system problem-solving approach from that of Geometry Tutor, where the system itself handles drawing automatically. So, students have the opportunity to demonstrate and entrench their creative skills, which is not available in Geometry Tutor, although this takes part of the learning time, which seems to be an advantage Geometry Tutor has over Andes.

Andes's student model is based on BN, which provides probabilistic estimates of a student's mental state. It combines the current state of the problem solving process with the long-term assessment of the student's knowledge in a probabilistic representation (Conati et al., 1997). With this, the system can take care of multiples sources of uncertainty due to the student's unconstrained learning actions pattern or order, beliefs, learning goals, and his/her domain knowledge level prior to the commencement of problem solving. The probabilistic estimates or representation guides help the decision-making process. Also, in pursuit of effective help in decision-making, the student model stores information about what problems the student has worked on, what interface features they used, and what help they have received so far from the system (Gertner & VanLehn, 2000). All the aforementioned assist in the computation of probabilistic estimates, which BN provides to Andes in order to determine the kind of help offered to the student (Gertner, Conati & VanLehn, 1998).

One positive aspect of Andes that has been identified, which provides educational benefit, is encouragement of constructive, as opposed to passive, learning (Gertner & VanLehn, 2000). This is achieved through feedback scaffolding at different levels, which encourages students to think. Scaffolding hints have been used in other intelligent tutors, and have been claimed to enhance learning from a problem-solving perspective (McKendree, 1990). The system can generate hints automatically to provide clues for the student to resolve an impasse, while the student can equally request help through the flexible help facility handled by the Andes procedural help unit. This may have its own disadvantage in that student may want to use it always, thereby preventing deep learning through thinking. However, the advantage seems to be greater in that it may eliminate frustration resulting from impasse during the learning activity.

Overall, Andes reflects the tutoring approach, feedback and hint generation techniques utilised to accomplish its underlying design principles, which cannot be said to be formally derived from a pedagogical theoretical framework. The ITS employs the model tracing approach for its tutoring strategy, visual immediate feedback approach for feedback on valid and invalid learning actions, and sequential hints for errors during the learning process. These strategies had been implemented in similar intelligent tutors, and were found effective (McKendree, Radlinski & Atwood, 1992; Anderson et al., 1995; Reiser et al., 2002). Despite the success of these strategies, the pedagogic implications of the immediate feedback and hint sequences strategies have come under criticism, which are captured thus:

- first, the system cannot detect shallow learning (Aleven, Koedinger & Cross, 1999)—a consequence of system provision which offers the student room to guess, until they arrive at positive feedback;
- second, the system does not request an explanation for student action—a form of qualitative reasoning (regarded as *"talking science"*), which has been found to

deepen learning and is a way towards understanding science (Van Heuvelen, 1991 cited in Gertner & VanLehn, 2000);

- third, the system does not allow stepping back to peruse previous steps the student took to arrive at the solution to a problem (Chi, Feltovich & Glaser, 1981; also cited in: Lee et al., 2010; Chi et al., 2011)—hence, neither the system nor student have the opportunity to trace back the solution path, which is possible in pencil-paper mode; and
- fourth, when learning takes place quantitatively, the system does not provide a qualitative or semantic perspective of the process—so, it fails to induce versions of skills required to solve qualitative problems, and to check quantitative work for reasonableness (Halloun & Hestenes, 1985 cited in Graesser et al., 2001).

It also lacks pedagogical theory-based formalisation, which could have informed a better design that takes care of some of the issues raised. Despite all these criticisms, on a fair ground, it can be said that there is no perfect system, and that the system has undergone a series of revisions that have been found to enhance learning, and are effective in improving problem solving skills. The criticisms actually open up research issues that should be considered to further improve ITS (e.g. apriori pedagogical theory based formalisation of ITS design).

2.6.3 Byzantium Intelligent Tutoring Tools

Byzantium (Patel & Kinshuk, 1997; Patel, Scott & Kinshuk, 2001) is a set of tutoring tools covering some topics in the domain of accounting, specifically, topics in introductory financial and management accounting. The central research issue addressed is how to simulate human tutoring capabilities in tutoring systems. Hence, it was developed with the aim of having a set of tools that can emulate a human tutor by providing students with a platform where they can learn, practise, and test their numerical skills.

According to Patel, Cook & Spencer (2003), the Byzantium project can be traced to the initial prototype software (i.e. marginal costing) developed at De Montfort University by Ashok Patel. Thereafter, Byzantium was developed by a consortium of six universities within the Teaching and Learning Technology Programme (TLTP) of the Higher Education Funding Councils of the United Kingdom between 1993 and 1997. The actual project hub was at De Montfort University, where Ashok Patel led a team of developers, Kinshuk, Jamie Hunter and Navjeet Megh in developing learning tools for financial accounting and management accounting, and the Byzantium Marker. This yielded a set of Intelligent Tutoring Tools (ITT).

In order to achieve the aim of the project, its development was driven by a set of objectives, carried over from the pre-Byzantium era prototype (i.e. initial marginal costing prototype) that first set them, which are: first, to enhance quality, speed and thoroughness of student learning; second, make more efficient use of teaching resources, particularly staff time deployed within formal teaching time, and for correcting student work and providing feedback; third, to make numeric disciplines more accessible to non-specialists; fourth, to encourage IT awareness and its integration in courses; and fifth, to improve access to educational opportunities by promoting learning tools suitable for flexible and distance learning delivery methods (Patel, Cook & Spencer, 2003). However, to achieve these objectives, it was expected that the tutoring tools should satisfy some performance standards (detailed in Patel, Cook & Spencer, 2003) in a way to determine their reliability and learning effectiveness. Accordingly, it was assumed that if the standards are met, the tools would go a long way to save the scarce and expensive staff time deployed in organising tutorial sessions for students; more so, classes are largest at the introductory stages, because students from various disciplines are expected to take introductory accounting courses.

Equally, it would afford students the opportunity to practice on their own without restriction of time and place, by employing the one-to-one tutoring strategy—an effective tutoring methodology (Corbett, Koedinger & Anderson, 1997) and the gold standard for measuring other methods of pedagogy (Bloom, 1984). However, it could be argued that the design standards—detailed in Patel, Cook & Spencer (2003)—were not initially informed by any theoretical frameworks, since they were not explicitly stated to have emerged from one, but were later linked to some theoretical elements (e.g. conversation; scaffolding, fading, etc.). Nevertheless, the success of the tutoring tools generated from the standards, which is discussed in Patel, Cook & Spencer (2003), does indicate possible greater success if the design was formalised right from the outset. This is the position upheld by this research, and if implemented, has the potential to improve other aspects of the tutoring tools, aside from feedback.

Each ITT employs a mixed-initiative approach (first suggested by Carbonell, 1970) and is driven by an applied inference engine that processes stored predefined domain rules. The inference engine is based on overlay architecture, also utilised or referenced by some studies (e.g., Liegle & Woo, 2000; Murray, 2003b; Moundridou & Virvou, 2003a; Mara, Rodrigues & Carvalho, 2004; Perez-Marin & Pascual-Nieto, 2010; Mitrovic, 2012). Structurally, all the ITT share the same inference engine, which is domain-independent, but each consists of four components, namely tutoring module, expert and knowledge module, student model and user interface, also known as communication module (Woolf, 2009). In addition, each ITT supports a level selector, which enables the selection of user level (student, lecturer or administrator).

The major functional elements of the ITT are the variables and the operators. The variables are empty containers connected in a network of inter-relationships (or what can be regarded as a neural connection of nodes), and accept any value entered, provided the whole network remains consistent (Patel & Kinshuk, 1997). These variables and/or operators combined are used to formulate rules that the inference engine requires to provide intelligent tutoring. Moreover, one vital feature, which enriches and stimulates learning, is the facility that generates questions, known as the random question generator. It allows an unlimited number of questions to be generated and presented to the student during learning based on question template(s) already defined by lecturer. So, each ITT does not need to have any data bank and each student can get individualised questions. Equally, each ITT has some pre-defined question template(s), which are modifiable any time by the lecturer as desired.

Byzantium ITT approach has provision for two types of feedback (immediate and delayed feedback), given according to system learning mode (interactive or assignment mode). Feedback can be a combination of an alert beep and text message or text message only. In *Interactive mode*, depicted as practice mode, each ITT provides the student with an environment for practising problem solving skills with guidance from the system via the feedback mechanism. While in *Assignment* mode, it allows students to test their knowledge of the domain, and so provides a delayed feedback, referred to as "static" feedback (as cited in Patel & Kinshuk, 1997). The implementation implication is that any values can be entered at random (no order or sequence restriction), then feedback is delayed until marking is carried out in a batch process. Routen (1992) argues, *"there are advantages with both forms of student monitoring. Static feedback perhaps is less obtrusive ... while dynamic feedback prevents students from making gross errors and getting completely lost"* (as quoted in Patel & Kinshuk, 1997). The Byzantium approach benefited from both feedback types through the implementation of the two assessment modes: the interactive and assignment modes (otherwise referred to as formative and summative modes respectively), each mode implementing one feedback type.

The Byzantium implementation of immediate feedback is in line with Anderson's Geometry tutor (Anderson, Boyle & Yost, op. cit) model-tracing tutoring paradigm, but differs in the manner of its implementation. While Anderson's implementation monitors steps constituting the process (or rule), the Byzantium approach monitors the student's outcome. Also, Matsuda & VanLehn (2005) provided an advanced version of Anderson's Geometry tutor by incorporating two types of feedback scaffolding in steps, namely proactive and reactive feedback. Each implementation approach has its merits and demerits, and each can be justified depending on the educational perspective of the argument. That said, evidence in the literature shows that students do have different learning styles (Pask, 1988; Sharples, 2002; Vasilyeva, Pechenizkiy & Puuronen, 2006; Kolmos & Holgaard, 2008; Alves, Pires & Amaral, 2009; Penger & Tekavcic, 2009; Popescu, 2009; Alseddiqi & Mishra, 2010; Hou et al., 2010). With this in mind, it can be argued that each feedback implementation may not meet the learning needs of all student users. Therefore, a system that accommodates both implementations will be quite good, thereby allowing the student to select the implementation that best stimulates and enhances deep learning, not obstructing thinking. Moreover, feedback is so crucial in a learning process (Melis, 2005; Shute, 2008) that its implementation should be well guided to achieve a high learning outcome, as well as enhance knowledge transfer.

Hence, this research intends to consider a new direction, through a pedagogic metamodel that could make available optional implementations, where the student will be given the opportunity to determine which one best fits his/her learning needs. Such an intelligent tutor could only emerge from a metamodel that provides authors (or lecturers) with rich theoretical elements. These elements should be optionally useable to build diverse ITSs, and then tested to evaluate their impact on students' misconception detection, feedback generation and learning effectiveness. Also, the implementation of such a metamodel should enable the extension, replication and enhancement of the features of Byzantium tutors, which up to now, has been impossible and requiring development from scratch, if they are to be advanced. It should also make available new ITSs for topics uncovered by existing Byzantium tutors.

Now that we have discussed some ITSs with issues/questions emerging, which need to be addressed, it will be fair to still acknowledge that the above discussion does not in any way provide a comprehensive account or overview of work in the ITS field. It only highlights salient issues that can form a good foundation for current research. This research cannot in any way discuss the entire ITS works undertaken to date due to the limitations on resources. In view of the foregoing, this work draws inspiration and new direction from the general review of the ITS field, and the selected works presented above, especially the Byzantium project, which happens to fall within the context/domain being considered in this research. The questions then are: how do we replicate and enhance the design of these varying tutoring tools in order to build new ones for other numerical topics uncovered? How can we achieve the modification and extension of these new tutoring tools, when necessary, without having to develop from scratch—the absence of which is known to have impeded further work on Byzantium tools since their creation? Or, does any work exist that addresses these questions, especially within the context and domain of this research? More importantly, how can we address these questions from a formalisation perspective that hinges on educational theories? Answers to all these questions require a review of authoring research and some theoretical perspectives, which are considered next.

2.7 Authoring Intelligent Tutoring Systems (ITSs)

Although, several ITSs have been developed for different domains—including numerical domains, very few ITSs exist in the numerical aspects of applied numerical domains (e.g. accounting, engineering) where this research is located. Generally, it has been claimed in the literature that some factors hinder availability and construction of useful ITSs (Koedinger et al., 2004; Blessing et al., 2009). These factors relate to ITS's nature and development; they include the fact that ITS construction is difficult and complex, time-consuming, expensive, and needs collaboration of experts in related fields (El-Sheikh & Sticklen, 1998; Virvou & Moundridou, 2001; Moundridou & Virvou, 2003a; Murray, 1999, 2003; Blessing et al., 2009; Woolf, 2009; Zarandi, Khademian & Minaei-Bidgoli, 2012). The factors impose a major bottleneck in ITS use (Murray, 1997); generally, they explain why very few reliable ITSs can be found in the classrooms (Virvou & Moundridou, 2001; Murray, 1997, 2003). Specifically, the factors could be attributed to the non-availability of more Byzantium-like ITTs (Kinshuk, Patel & Russell, 2000), after the first four ITTs were successfully developed and evaluated in the accounting and finance domain (Patel, Cook & Spencer, 2003). Thus, the need for an ITS authoring tool arises that can bridge this gap or overcome the problems demonstrated.

A survey of the literature shows that commercial authoring tools are available. However, they are best suited for traditional computer aided instruction (CAI) systems and multimedia-based training, because they lack the sophistication required to construct intelligent tutors (Murray, 1999, 2003a, 2003b). These commercial authoring systems excel by providing rich tools that the instructional designer can utilise to produce visually-appealing and interactive screens; however, behind these screens you find a shallow representation of content and pedagogy strategies (Murray, 2003a). This calls for specialised authoring tools that can address this weaknesses and overcome the problems associated with ITS construction. Accordingly, it is worth noting that some efforts have been channelled towards ITS authoring research. These efforts yielded some results, leading to the development of varying ITS authoring tools that attempted to make ITS construction easier; see Murray, Blessing & Ainsworth (2003), for a detailed review.

The resulting authoring works attempted to enhance construction of cost-effective ITSs (Murray, 2003b), and have been classified into either authoring tools or shells. According to Murray (2003a), an authoring shell is a generalised framework for building ITSs, while an authoring tool is an ITS shell along with a user interface that can be utilised by non-programmers to formalise and visualise their knowledge. In this research, the perspective of an ITS authoring tool was adopted since this work aimed the formalisation of the design of an authoring tool that can produce an inventory of ITSs for the applied numerical disciplines. Also, the perspective provided a platform to test the implementation of a formal (theory based) approach in an ITS authoring tool.

Furthermore, there are claims that over two dozen ITS authoring tools/shells have been developed (Murray, 1999, 2003a). These include Eon (Murray, 1998, 2003a), REDEEM (Ainsworth et al., 2003), IRIS (Arruarte et al., 2003), CREAM-TOOLS (Nkambou et al., 2003), and WEAR (Virvou & Moundridou, 2000; Moundridou & Virvou, 2003a), etc. Notwithstanding these early works, new ITS authoring tools/shells are still emerging, pushing up the number of available tools; e.g. Fuzzy-based framework (Zarandi, Khademian & Minaei-Bidgoli, 2012), CTAT (Blessing et al., 2009; Aleven et al., 2006a, 2006b; Koedinger et al., 2004), xPST (Gibert, Devasani & Kodavali, 2011), ASPIRE (Mitrovic, Martin & Suraweera, 2009), SMARTIES (Hayashi, Bourdeau &

Mizoguchi, 2009), and LEMONADE (Giemza et al., 2010), etc. The new tools thus indicate the significance and currency of research in the ITS authoring field. An overview of these tools shows that they were primarily designed to generate ITSs within a short space of time, and to eliminate the developmental expertise required, consequently leading to reduction of cost and time associated with the development of individual ITS.

The above-mentioned authoring works employed diverse strategies and approaches to achieve their design goals. These goals included the usability of an ITS authoring tool and reusability of its products due to their significance (Virvou & Moundridou, 2000; Brusilovsky, 2003; Blessing et al., 2009; Gilbert et al., 2011); otherwise, their purpose would be defeated. Blessing et al. (2009, p.196), in an attempt to express the significance of usability, stated:

> *A major risk associated with this project was not that the resulting tools would not author meaningful cognitive tutors, but rather that the tools would be too complex for authors who are not cognitive scientists or programmers to understand.*

On the other hand, Brusilovsky (2003, p.403) noted the importance of reusability as follows, "..*In the near future we should expect more powerful authoring systems and frameworks that combine adaptive hypermedia and courseware reusability ideas.*"

Existing ITS authoring tools are also associated with different levels of sophistication, in terms of ITS components construction. Some concentrated on the construction of a few ITS modules, such as the construction of domain and student modules; for example, ASPIRE (Mitrovic, Martin & Suraweera, 2009), and CTAT (Blessing et al., 2009). There are also a few tools that addressed the conventional four components of an ITS, e.g. EON (Murray, 1998, 2003a). Some of these tools can only construct courseware ITSs, some allow the construction of tutorial-like ITS, and some do have capabilities for both. However, the research discussed in this thesis adopted the basic four-component structure (as argued in section 2.5) required in tutorial-like ITSs. More so, each of the components would play a significant role in ITSs that are meant to learn numerical aspects of applied numerical disciplines. Thus, this research focused on an ITS authoring tool that enabled the construction of the domain, tutoring and interface modules; also, it provided a generic algorithm that can manage the student modelling aspect of the constructed ITS.

In realisation of the variations in ITS types that can be authored, Murray (1999, 2003a) further classified ITS authoring tools into two broad categories, based on their capabilities or ITSs they produce. Thus, pedagogy-oriented tools were those ITS authoring tools that focus on how to sequence and teach canned domain content, while performance-oriented authoring tools were those that focus on providing a rich learning environment, which supports learning by practise and receiving feedback (Murray, 2003a). Nkambou, Bourdeau & Psyche (2010) added a third category, which they referred to as instructional-design-oriented tools. In the light of these categories, current research adopts a performance-oriented ITS authoring tool, since by nature, performance-oriented learning environment involves engagement in activities which match the learning requirement of this research's context (numerical problem solving).

Also, the foregoing was informed by the theoretical frameworks intended for the design of the ITS authoring tool (see section 2.9 below), which assumed that knowledge exists but is advanced through construction with support from at least a domain expert. In order to "operationalise" the assumption, this research assumed learners would have learnt declarative aspects of knowledge through traditional classroom-based approach, reading, etc., thereby having prior knowledge of the domain of interest. What would then be needed is to support classroom teachers through provision of a tutoring system that takes care of tutorial sessions where procedural knowledge—involving practising—can be learnt. So, an authoring tool that can produce intelligent tutors with an environment that supports "*learning by doing*" was considered appropriate, and falls within the category of performance-oriented tools. At this point, one may be tempted to ask: are there any appropriate and formalised ITS authoring tools for the context/domain envisaged in this research?

With a view to addressing the latter question, it should be noted that despite achievements in authoring research, many of the problems associated with the construction of useful ITSs still remain, even though some are still unexplored and have been identified in Murray (1999, 2003a). However, of interest is that none has been designed for and/or extensively evaluated in the numerical problem-solving context of applied numerical domains involving categorisation and application of rules (e.g accounting)—the contextual/evaluation focus of this research. There is a realisation that research could be treated in context and still produce extendable results (Nardi, 1996 cited in Luckin, 2010). Also, on realising that a key factor that could contribute to the successful implementation of an ITS authoring tool is to limit the extent of domain generalisation.

The foregoing arguments could be sustained, considering Hsieh & Hsieh's (2001) justification for the development of XAIDA—a simulation-based ITS authoring tool for the manufacturing domain. They argued: "*....development in the area of manufacturing engineering has been rare. However, recent developments in the area of ITS authoring tools may make this technology more accessible.*" (Hsieh & Hsieh, 2001, p.569). This argument also aligns with other views in the literature. For instance, Virvou & Moundridou (2000) claimed that developing an ITS authoring tool for all domains seems unfeasible, while Murray argued that to achieve a usable and powerful ITS authoring tool, its development should be based on some underlying formalism that "*....will satisfy the needs of some types of tutors yet not be appropriate for authoring other tutors*" (Murray, 2003a, p. 515). This view thus portends that ITS authoring tools may not be applicable in all domains. Furthermore, this stance could be entrenched considering another research effort that addressed the development of a generic ITS authoring tool that enables easy creation of ITSs for multiple military synthetic environments (SEs) (Gilbert et al., 2011). This work demonstrates that generalisation is limited to the synthetic environments only, confirming earlier stated views.

On that note, one cannot guarantee the implementation of existing ITS authoring tools to generate meaningful intelligent tutors in the applied numerical problem-solving context/domains. More so, the current research could not substantiate availability of such tools—for now—in the context/domain under consideration. Also, if ITS development should be formalised as earlier argued in previous sections, corresponding authoring tools should be formalised too. This view aligns with Murray (1999, 2003a) who claimed that ITS authoring tools should employ customisable formalisation. However, the formalisation considered in this work envisaged the implementation of educational theories, since ITSs are educational systems (Conati, 2009) and ITS authoring tools are meant to generate usable and meaningful intelligent tutors for educational purposes.

However, a known trend is that many ITS authoring tool designs seem driven by two main disciplines—AI and cognitive science—that have been dominating ITS/Authoring research, right from its inception. So, what you find, in most cases, is the implementation/testing of AI methodologies (e.g. Zarandi, Khademian & Minaei-

Bidgoli, 2012), and implementation of methodologies that emerged from psychological models, such as ACT-R cognition model, e.g. Cognitive Tutor Authoring Tools or CTAT (Blessing et al., 2009; Aleven et al., 2006a; Aleven et al., 2006b) and constraint-based models, e.g. ASPIRE (Mitrovic, Martin & Suraweera, 2009). Additionally, the field of human computer interaction (HCI) and existing commercial authoring tools also impacted the design of ITS authoring tools. This led to intuitive design and/or enhancement of commercial authoring tools features in ITS authoring tools, as claimed in the design of EON—a pedagogy-oriented ITS authoring tool (Murray, 1998, 2003b). These examples further throw light into where ITS authoring tools derived their design from, predominantly AI, cognitive science, and intuitive/HCI-oriented design. Also, although the pedagogy-oriented tools (e.g. EON, REDEEM, etc.) do not fall within the ambit of this research, it could be concluded that there are tools that support performance-oriented features—the focus of this research, such as, for example, CTAT (Blessing et al., 2009), and WEAR (Moundridou & Virvou, 2003a). However, the applicability of the latter in the context/domain and design underpinned by educational theoretical frameworks, as intended in this research, cannot be ascertained, since there is no evidence to that effect.

The above further entrench Self's (1990b) argument, that most ITS designs were not shaped by formal theory, or at best, linked to informal theories. While the claim applies to ITSs, it is reflected in ITS authoring works too. As a result, the need to bridge the gap between theory and practice arises, which Hartley (2010) established with respect to CAI systems. That said, there seem to be few exceptions—with respect to ITS authoring tool—traceable to the literature, in terms of their design derivation that emerged from the educational theories. These ITS authoring tools are: CREAM-Tools or Curriculum REpresentation and Acquisition Model-Tools (Nkambou, Frasson & Gauthier, 2003) and SMARTIES (Hayashi, Bourdeau & Mizoguchi, 2009). CREAM-Tools is a set of tools that allows creation and organisation of a curriculum according to the three models: the domain, the pedagogy and the didactic (i.e. resources) aspects. The authoring tools design was informed by Gagne's taxonomy (cited in Nkambou, Frasson & Gauthier, 2003)—an hierarchy or classification of types of knowledge or learned capabilities (intellectual skills, cognitive strategies, verbal information, attitudes, and motor skills) required to achieve learning. These tools were designed for specific domains, such as simulation and the teaching of concepts (ibid). As such, it was classified to be more of a pedagogy-oriented tool (Murray, 2003a), and claimed to have some elements of performance-oriented features, which has not been evaluated or proven empirically. Being a simulation-based and pedagogy-oriented ITS authoring tool, it falls outside the scope of this research, which is about a performance-oriented ITS authoring tool. Also, Gagne's taxonomy was only utilised to inform KR. This contrast with the approach is intended in this research, in which all aspects of the authoring tool are informed by the theory-based formalisation.

One other hand, SMARTIES (Hayashi, Bourdeau & Mizoguchi, 2009) is a theory-aware authoring tool that is linked to an ontology named OMNIBUS (Mizoguchi, Hayashi & Bourdeau, 2010). It uses multiple learning/instructional theories, adopts a theory-ontology engineering approach, and applies AI agents in its implementation. This authoring work was driven by the necessity to answer the call of Self (1990b) with respect to theory-based formalisation. This is a call that is considered genuine and due, as acknowledged by Nkambou, Bourdeau & Mizoguchi (2010), and in which SMARTIES is seen to be a step towards honouring. However, SMARTIES lacks specific pedagogical focus due to its multiple theories (i.e., it utilised nine theories), therefore requiring the AI agent to infer the author's pedagogy. Unlike SMARTIES, this research has the advantage of having a pedagogical focus, using two constructivist theoretical frameworks, thus, not requiring AI-agents to infer author's pedagogy. Also, SMARTIES falls under the pedagogy-oriented category, and as such, does not fit into the context of this research, since it cannot generate ITSs implementing the conversation-cognitive approach. It has also been criticised and acknowledged to be complex (Hayashi, Bourdeau & Mizoguchi, 2009). This makes it unsuitable for its eventual users, who might be educationalists and non-programmers, hence defeating its purpose. Therefore, its suitability for domain experts (who are non-programmers) in the applied numerical problem solving context/domains (e.g. accounting)—where current work is evaluated—is questionable, since it is a pedagogy-oriented authoring tool which falls outside the ambit of current research; more so that Hayashi, Bourdeau & Mizoguchi (ibid) acknowledged that their evaluation of the tool shows it is far from practical to use.

Also, other existing ITS authoring tools that have a simple approach may not be suitable either, because, as argued earlier, if these tools are meant for educational purposes, conventional educational theories should play a significant role in shaping their design. However, they lack an education-driven theoretical foundation, which current research intends. More so, the product of any ITS authoring tool tends to reflect the latter's underlying assumptions or notions. So, an ITS authoring tool underpinned by pedagogical assumptions drawn from conventional teaching and learning theories could adequately reflect the pedagogical strategies and elements of its underlying theories. Such an educational approach to design could clarify a tool's ontological and epistemological background, which may be helpful in furthering and understanding research based on such a tool. It has the advantage of predicting the possible learning process of its product and can also help in decision-making, when choosing among ITS authoring tools to support the traditional teaching and learning process. The latter point is critical, in that a purchaser of an instructional support tool might ask: "*what is really available (or soon to be available) to make ITS authoring cost effective?*" (Murray, 2003a, p.492). It also could help tailor ITS construction towards achieving teaching and learning goals supported by the parent authoring tool, instead of the trend in the literature in which new AI implementations or cognitive computational models are tested or simulated.

Furthermore, formalisation of ITS authoring tools using educational theories could benefit from the rich theoretical foundation of the education field. It could inform a good design since many of the conventional educational theories, e.g. CT (Pask, 1996), and CA theory (Collins, Brown & Newman, 1989; Collins, 1991a) do capture various teaching-learning variables that are applicable in a technology-enhanced learning environment. For example, theories could: inform appropriate learning context/settings, identify features that enhance and/or inhibit teaching-learning process, suggest appropriate KR, and inform learning content/feedback scaffolding etc. Theories stand to be valuable conceptual frameworks for the design of useful ITS authoring tools. More so, that some elements of these educational theories (e.g. scaffolding, fading etc.) have been mentioned in some ITS/Authoring tools (indicating their significance) but without the formal implementation of these theories, e.g. WEAR (Virvou & Moundridou, 2000), and Advanced Geometry Tutor (Matsuda & VanLehn, 2005). The question is, if these conceptual variables are valuable to these tools, then why not formalise the design of these tools using educational theories, since the variables emerged from these pedagogical theories? This is an open question that queries the intent behind their use in these tools. This informed the need to formalise ITS authoring tools' design/development using educational theories. In order to be relevant in this research's context/domain, such formalisation should be apriori and implemented in the applied numerical problem-solving context/domain, to generate relevant and useful ITSs. This provided a window to test the applicability of an educationally-driven, theory-based design in the stated context and domain involving categorisation and application of rules.

2.8 The Thesis Focus

The above review of ITS/Authoring research projects many possible investigative issues that cut across different aspects of the field, and in which some were identified in the discussion. In order to focus this research, it draws on a key motivating factor of ITS research, the implementation of the human one-to-one tutoring strategy that has been considered an effective teaching and learning strategy (Bloom, 1984), and which has also been acknowledged in several works such as Koedinger & Corbett (2006), Koedinger & Aleven (2007), Lane & Johnson (2008), Chi & VanLehn (2010), and Chi et al. (2011), resulting in many ITS implementations. Based on the educational origin of this strategy, this research argues in favour of educational theories shaping the design of ITSs, since they are meant for educational purposes (see: Ainsworth et al., 2003; Conati, 2009; Chaouachi et al., 2010).

Also, the need for ITS authoring research that is shaped by educational theories was identified in section 2.7 (above). The above review identified the trend in ITS authoring research which also showed that many designs were not formally linked to theories, a trend that was also recently acknowledged in Nkambou, Bourdeau & Mizoguchi (2010). This trend could be attributed to the drive to exploit ITS—AI/psychological alignment (e.g. Ainsworth et al., 2003; Conati, 2009), which could be connected to the foundation of many of the leading researchers in the field (principally from an AI and cognitive psychology background). This stance was reflected in the argument of Self (1990b)—an AI researcher, when he called for the formalisation of ITS design to emerge from AI. However, one area of interest that seems not exhaustively exploited is the pedagogy engineering of ITS authoring tool design, which is AI-, or cognitive science-neutral, a design approach that is underpinned and draws on the significant role educational theories could play in ITS.

Aside from the above, one key success of ITS is the relevant and timely guidance/feedback a student receives during learning. Its significance is further reinforced by Shute (2008, p.1), in that:

> *Feedback used in educational contexts is generally regarded as crucial to improving knowledge and skill acquisition........ In addition to its influence on achievement, feedback is also depicted as a significant factor in motivating learning........ However, for learning, the story on feedback is not quite so rosy or simple.*

As a result, feedback is considered a vital element that should be considered in the design of an ITS authoring tool, to achieve production of meaningful ITSs. Although some ITS research has examined feedback, it is still open to further and well-grounded research. Thus, its consideration in the context of the theory-based formalisation is seen as vital; otherwise, the emerging ITS authoring tool might defeat its usefulness, if ITSs generated cannot provide meaningful feedback. Also, this enabled the investigation of the role of theoretical constructs (e.g. cognitive visibility) in feedback generation.

It has also been argued that human tutors give two types of feedback in response to learning actions: one, to correct errors during learning; and two, a reward for the learner's action. Accordingly, such feedback has been referred to as negative and positive respectively (Ohlsson, 1996b; Koedinger & Aleven, 2007; Mitrovic, 2012). Furthermore, Ohlsson (1996b) argued that learning involves two cognitive functions, error detection and error correction. The former could be referred to as misconception detection and the latter feedback (i.e. negative feedback). However, the latter also admitted that people do not solely learn by correcting their errors, they could benefit from positive feedback. The foregoing thus suggests a connection between feedback and misconception. So, investigating both constructs within the theory-based formalisation perspective is considered viable and novel, since this approach cannot be traced to the literature. In that sense, it would provide a platform to examine the impact of the ontological and epistemological assumptions of the theoretical frameworks as implemented in the ITS authoring tool designed in this research with respect to the two constructs.

Moreover, determining appropriate theoretical frameworks, this research takes into account some key factors, which are likely to have direct and indirect impact on it. These include: the nature of applied numerical disciplines and related problem-solving techniques, and the ways and manner domain experts engage students during tutorial sessions. According to Patel, Scott & Kinshuk (2001), numerical domains are second order in nature, because reality is represented as numeric models that are manipulated to yield results. As such, domain experts work with concepts (e.g. volume, force, kilometres, monies, etc.) that are abstracted to a numeric representation. For instance, accountants deal with social and economic realities that are expressed in monetary terms, while engineers deal with abstract mathematical measures such as force and gravity. Thus, learning such domains does involve problem solving and calls for greater utilisation of cognitive skills, which makes them more difficult to grasp (ibid).

Pask (1975) argued that learning in any domain entails understanding the relationships between its concepts. When contextualised and concretised, this translates to working with numeric variables, which represents domain concepts. So, learners need to understand the interrelations that exist among them (Patel, Scott & Kinshuk, 2001), to distinguish between independent and dependent variables, and know how to manipulate known variables to discover unknown variables, in attempts to achieve the overall learning goal. This therefore involves hands-on experience or learning-by-doing, which can be regarded as a problem solving technique. In the same light, Anderson, Boyle & Yost (1985) argued that learning numerical domains (e.g. geometry), requires gaining mastery and competency that can be acquired through practising of problems. This involves a series of well-sequenced elementary operations that when clustered, could constitute the procedural knowledge of a domain. Patel, Scott & Kinshuk (2001) argued that this procedural knowledge distinguishes a novice from a domain expert. As such, the domain expert needs to scaffold the procedures in a logical and step-wise manner for a novice to learn, and this involves conversations or information exchange between the two. So, in terms of technology-based tutoring systems, such domains seem to favour CA in a conversational environment.

Furthermore, this research acknowledged the enormous cognitive tasks involved in knowledge construction during problem solving. It aligns with Ohlsson's (1996b) argument that two learning functions, error detection and error correction, constitute part of cognitive activities that play-out in an attempt to construct knowledge. In the same light, Collins, Brown & Holum (1991) argued that if students' cognitive process could be made visible during learning, the master (domain expert) would be in a position to provide relevant feedback. So, to diagnose a student's problem-solving difficulty, domain experts looked into problem-solving steps, which can be said to be a representation of the cognitive task of a learner. From the foregoing, two key issues emerge: tutoring through conversation, and cognitive visibility of the student's problem-solving process. Hence, to provide adequate and appropriate guidance that is effective in one-to-one tutoring, a conversational approach that takes into consideration the student's cognitive problem-solving process, through task-steps identification, was considered essential.

Moreover, the research noted that in a conventional teaching environment, the human tutor-student conversation that takes place in an attempt to achieve learning constitutes the basic assumption that Gordon Pask embraced in CT (Pask, 1976a, 1976b, 1976c, 1988; Scott, 2001a; Boyd, 2004; Sharples, 2005; Heinze, Procter & Scott, 2007; Heinze & Heinze, 2009). Likewise, learning can be achieved when activities are situated and involve two or more participants with one being advanced in the target domain, a position captured in CA (Collins, Brown & Newman, 1989; Brown, Collins & Duguid, 1989; Dennen, 2004; Dennen & Burner, 2008). Also, guidance can be provided by a human tutor, if learners' misconception or missing conception could be identified, or if learners' cognitive processes can be made visible (Collins, Brown & Holum, 1991). The foregoing thus highlights two key theoretical concepts (i.e. conversation and cognitive visibility) that have been part of human tutoring strategies, even when they are undertaken unconsciously, their consideration in technology-driven tutoring systems seems viable. Therefore, one research issue that can be drawn from these two theories is how conversation can be used to enhance the visibility of student's cognitive process.

However, contextualisation of the theoretical concepts (i.e. conversation and cognitive visibility) would be necessary to enhance their applicability in this research. Conversations mean non-verbal information exchange or dialogue between the user and the system in the form of interactive engagement and bi-directional communication that advances learning. On the other hand, cognitive visibility refers to the conceptual visibility of minute cognitive steps undertaken during a problem-solving endeavour. These two concepts were explored through an approach known as process monitoring (i.e. PM), used as an intervention to achieve improved cognitive visibility. Thus, PM augments the conversation within a cognitive apprenticeship framework, implemented as one of the tutoring strategies that an ITS authoring tool provides. The approach was based on the assumption that a human tutor provides relevant guidance to a student during one-to-one conversation, if the student's cognitive process is made visible. Thus, ITSs should also be able to mimic this style. However, to enhance the availability of such tutoring systems in the context/domain under consideration, an ITS authoring tool was developed and its implementation/development was formalised using the theories identified above.

This research explored the design of an ITS authoring tool from a pedagogic metamodel that is based on the synthesis of CT and CA theories. Despite their combined exploration, the research acknowledged previous application of the theories in varying contexts, e.g. CT was used in a blended learning study (Heinze, Procter & Scott, 2007, Heinze & Heinze, 2009), and CA was referenced in an ITS plan recognition study (Conati & VanLehn, 1996), while CA was applied in a Web-based study (Dickey, 2008), and CA drove a collaborative learning ontology (Isotani et al., 2009). Also, the implementation of some theoretical elements of CT (e.g. conversation) and CA (e.g. scaffolding and fading) were claimed in Byzantium ITT (Patel, Scott & Kinshuk, 2001). However in this research, a comprehensive view of the theories is being considered with interest in how conversation and cognitive visibility, implemented as PM, impacts detection of misconception, generation of timely and relevant feedback, and in turn enhances learning effectiveness. Also, the above stated theoretical concepts were considered from the perspective of authoring research, unlike the above-stated ITS-based studies. The aim was to enhance the generation of intelligent tutors that support both conversation and cognitive visibility within the numerical problem-solving context of applied numerical domains.

Additionally, the approach enabled the implementation of a multiple tutoring strategies environment, in which process monitoring and the commonly-known model-tracing approach were supported by an ITS authoring tool. The design enabled authors (i.e. lecturers, the domain experts that are non-programmers) build ITSs with either strategies or both. It also provided students with flexibility, such that they can choose the tutoring route that suits their needs within a multi-tutoring strategy ITS. Thus, the formalisation of an ITS authoring tool, underpinned by a pedagogic metamodel using both CT and CA theories, and subsequent production of ITSs in the context/domain under consideration, seemed viable and novel; moreover, that the implementation of this approach—using both theories—in an authoring-based research was not traced to the literature.

The research output—a pedagogic metamodel-based ITS authoring tool—has some potential benefits, theoretically and practically. Conceptually, the work contributed to knowledge in the field of technology in education, through the conception of a metamodel based on the augmentation of conversation in a cognitive apprenticeship framework. It provided a window to investigate and understand the theoretical foundation, strategies and elements captured in the metamodel. In a practical sense, it stands to benefit institutions (as corporate bodies) as well as individuals (teachers and lecturers). Institutionally, the need for programmers to develop ITSs required to support teaching and learning will be eliminated—a way to save cost and drastically reduce time involved in developing new ITSs and the maintenance of old ones. Institutions can then properly deploy scarce resources to other areas of need.

Individually, domain experts (i.e. lecturers) can easily create tutoring tools to support their teaching activities, eliminating expert programmers required to build each ITS from scratch, thus, maximising their limited resources and channelling their energy to other pedagogy activities. This would have been impossible, if they had to provide one-to-one tutoring to a large student population, in the absence of such an authoring environment. The provision of dual-mode in ITSs constructed would support both Web and desktop features, giving students the flexibility in terms of learning modes—either online or desktop. It has the potential to reduce the learning curve inherent in the learning of numerical disciplines, since students will have direct access to one-to-one tutoring via technology, a service that could have been provided by the domain expert, but which may not be feasible due to a large student population and limited resources.

2.9 The Research Theoretical Foundation

In the light of the above review, the current research aims to derive a theoretical foundation from the discipline of education. It considered two conventional educational theories to develop a pedagogic metamodel that underpins the design of a prototype ITS authoring tool, known as Intelligent Learning Activity Builder System (ILABS) in this research. The theories consist of CT (Pask, 1996) and CA theory (Collins, Brown & Newman, 1989; Collins, Brown & Holum, 1991), as justified above.

CT is a theory of learning and teaching (Scott & Cong, 2010) propounded by Gordon Pask (Pask, 1975). It originated from a cybernetic and dialectic framework (Scott, 2001b; Sharples, 2005). Thus, it was considered a cybernetic theory but has found real world use in education (Scott, 2007). CT offers to elucidate how interactions between participants in a conversation lead to the "construction of knowledge" (Scott & Cong, 2008), promoting a "radical constructivist" epistemology of human learning (Scott, 2001b). Participants are 'psychological (p-) individuals', coherent conceptual systems, embodied in ‘mechanical (m-) individuals’ (brains, bodies and augmentations) (Scott & Cong, 2008). In that regard, CT distinguishes two sorts of stable and self-reproducing systems, namely the P-individuals and M-individuals. Pask describes conversation as a *P*-individual (a self-reproducing class of procedures) that is executable in one or more restricted class of *M*-individuals (processors, brains, bodies and augmentations) (Pask, Scott & Kallikourdis, 1973; Boyd, 2004). Thus, the theory emphasises conversations as media of knowledge construction and the need for a “knower” (student) within the communication space. Also, the theory recognises the existence of distinct domains of knowledge and the distinction between these domains are subject to negotiation and agreement within the conversations, constituting a community of observers (i.e. participants/individuals in a conversation). Equally, CT put forward a methodology―knowledge and task analysis―for analysising the structure of different knowledge domains. Therefore, as a whole theory, CT is a framework for understanding knowledge and reality, and it included the epistemology and methodology for investigating the world of objective reality. For a detail account of the theory, see chapter 3―section 3.2―of this thesis.

On the other hand, CA is regarded as a socio-constructivist approach to human learning, which advocates a master-apprentice relationship in a situated context for successful learning (Collins, Brown & Newman, 1989; Brown, Collins & Duguid, 1989; Collins, Brown & Holum, 1991; Dennen, 2004; Dennen & Burner, 2008). The theory emphasises the role of cognition in the learning processs, the social context of learning and the need for learning to commence from activity to abstraction. Unlike traditional apprenticeship―in which learning is external and bears a concrete product, cognitive skill is hidden (or internal) and needs to be open so that participants in the social learning process can support each other in constructing knowledge. So, CA approach aims to open-up the tacit cognitive processes so that the learner can observe, enact, and practice them with help from others in the social learning space (Collins, Brown & Holum, 1991). In this thesis, the process of achieving the foregoing is considered as cognitive visibility. In order to open-up the tacit processes running in both learner and master, the theory suggested six methods (modelling, coaching, scaffolding, articulation, reflection and exploration―see chapter 3 for their meaning. Also, the chapter provided a detailed description of the theory).

Although, there are other relevant theories, the choice of CT and CA stems from the tenets upon which these theories are founded (i.e. conversation and cognitive visibility), which constitute key elements and favour the context of this research, as argued above. Both theories emphasises the construction of knowledge in order to achieve learning. Also, their choice aims not only to clarify the theoretical and ideological assumptions that underpin this research, but also to offer a basis for a coherent argument and investigation of the research constructs identified earlier. The above thus represent insights into the theories (i.e. CT & CA), while chapter 3 provides a detail discussion, leading to the conception of a pedagogic metamodel that underpins the research discussed in this thesis.

2.10 Emergence of the Research Questions

Considering the above review, ITS/Authoring research issues, and the methodology for investigating them, may be said to be biased towards AI rather than education, although education forms the testing platform for these various ideas (e.g. Ben Ammar et al., 2010; Johnson, 2010; Baker, Goldstein & Heffernan, 2011; Mikic-Fonte et al., 2012; Zarandi, Khademian & Minaei-Bidgoli, 2012). The rise of cognitive science, which also contributed enormously to the ITS/Authoring field, compounded it, bringing to the fore different views on the type of research questions that should be investigated. While some views addressed the psychological aspects of learning—affect, emotion, motivation, etc. (e.g. Graesser & D'Mello, 2012), some examined achievements, effectiveness, KR, feedback impact, modelling of student behaviour, plan recognition, among others (e.g. Corbett & Anderson, 2001; Liu et al., 2011; VanLehn, 2011).

Notwithstanding the various investigations that seem connected to education, one recurring theme—as inferred from Graesser, Conley & Olney's (2012) definition of ITS—is that these investigations are driven by attempts to implement computational models from cognitive sciences, AI etc. For example, Baker, Goldstein & Heffernan (2011) assessed the probability that a student learned a knowledge component through a machine-learning model developed. Graesser, Conley & Olney (2012) studied the affect issue through application using a signal processing concept. Although these works were undertaken in an educational context, the drive was actually to test the effectiveness of their approaches. On that note, the argument of Koschmann (2001) seems relevant. He argued that research conducted under the ITS paradigm raises a set of questions whose focus is different from that raised by the CAI paradigm. This could be explained by the AI and cognitive science disciplines that significantly drive ITS studies, unlike CAI that emerged without AI or cognitive science influence. While CAI reconfigures education through strong links with theories via technology (Hartley, 2010), the majority of ITS researches were driven by AI or cognitive principles, which were tested in education.

Consequently, Koschmann (2001) claimed that CAI researches focus on instructional efficacy, while ITS researchers tend to address instructional competency (of implemented techniques). Hence, AI researchers are more interested in the fidelity of the tutoring system's performance (an AI objective), rather than its effectiveness in terms of students' learning outcome (an educational objective). The latter claimed that the shift in priorities accounted for the misunderstanding among researchers working within the two paradigms. Inasmuch that we do not query the validity of the argument, this research holds that both focuses are achievable within the ITS paradigm, regardless of the influence of AI and cognitive science. Both priorities are important, and from the educational point of view, they should be met within any tutoring system, if they are created purposely for education. In furtherance to that position and as earlier argued, education—as a discipline—should greatly inform the design of these systems. This position draws heavily from various arguments that have been put forward in this chapter. These included the need to formalise ITS design as argued in Self (1990b), the interplay between theory and practice (Hartley, 2010), and recent researches on tutoring system effectiveness and other related educational issues that have been investigated (e.g. VanLehn, 2011).

The research position, therefore, provides the foundation upon which it intends to investigate the theory-based formalisation of ILABS design, through the conception of an education-driven metamodel. ILABS should support the basic assumptions of such a metamodel (e.g. conversation, cognitive visibility etc.). The ITS authoring tool should allow authors (i.e. lecturers) to construct an inventory of intelligent tutors that support cognitive visibility of students' learning pattern during the learning process, being part of the issues being explored in this research. Also, ILABS should enable authors to optionally construct ITSs with multiple tutoring strategies. Such tutoring implementations allow personalisation by students. The students can decide which route to follow depending on which will enhance their learning outcome. Authors (i.e. lecturers) can also restrict students ..to a specific tutoring approach depending on what they intend to achieve by such an action. This is a clear distinction from the single tutoring approach implemented in the Byzantium project (see Patel & Kinshuk, 1997; Patel, Scott & Kinshuk, 2001), and other ITS researches.

Based on the totality of the review undertaken in this chapter and the research objectives—to conceptualise a pedagogic metamodel that can be implemented, assess usability of ILABS (an implementation of the metamodel), and explore tutoring strategies for cognitive visibility of student's learning process—this thesis aims to address the following research issues: [i] the conception of a pedagogic metamodel that is underpinned by conventional educational theories; [ii] formalise the design of an ITS authoring tool, and thereafter construct ITSs through implementation of the metamodel; [iii] support cognitive visibility of the learning process in ITSs constructed (a basic assumption of the metamodel); [iv] support optional multiple tutoring strategies as part of the features of the metamodel; and [v] evaluate the ITS authoring tool and its products (ITSs), in order to validate their alignment to the theoretical assumptions of the metamodel that underpins them, as well as assess the usability of the ITS authoring tool, especially for non-programmers in the accounting and finance domain. Hence, two research questions were posed and each was further broken down into four propositions as stated below:

Question One: What is the perception of users on the conception and implementation of a pedagogic metamodel in ILABS, which can be utilised by non-programmers to generate an unrestricted number of tutoring systems in a numerical problem-solving context of applied numerical domains?

- ***Proposition 1.1:*** A pedagogic metamodel can be conceptualised and implemented in ILABS for the applied numerical problem solving context.
- ***Proposition 1.2:*** It is possible to generate tutoring systems that support process monitoring and model-tracing from the implemented metamodel (i.e. ILABS).
- ***Proposition 1.3:*** The implemented metamodel can be used to create an unrestricted number of tutoring systems within a short space of time by authors (i.e. lecturers) who are non-programmers.
- ***Proposition 1.4:*** Users of the implemented metamodel have a positive perception about its ease of use and usability.

Question Two: Can the learner's cognitive process be made visible to aid the generation of relevant and timely diagnostic feedback in order to enhance learning effectiveness in the numerical problem-solving context?

- ***Proposition 2.1:*** The learner's cognitive process can be made visible to the tutoring system (or domain expert).
- ***Proposition 2.2:*** Cognitive visibility can be used to aid the generation of relevant and timely diagnostic feedback.
- ***Proposition 2.3.*** Cognitive visibility exposes/tracks learner's misconceptions.
- ***Proposition 2.4:*** Cognitive visibility enhances learning effectiveness.

In order to answer the above questions, chapter 3 addressed the conception of the metamodel and its implementation in ILABS, chapter 4 provided the evaluation methodological basis of the empirical studies undertaken. Chapters 5 and 6 provided the findings from the empirical evaluation of the implemented metamodel to determine alignment with the theoretical assumptions of the model and other issues in the above questions.

2.11 Summary

The review undertaken in this chapter explained many issues germane to the tutoring systems field in general, and to the ITS/Authoring research in particular. The work considered how the ITS field evolved from its precursor—CAI systems, in an attempt to introduce the concept of "intelligence" in tutoring systems. The attempt was to deliver systems that could address problems identified in CAI systems and to enhance learning. However the construction of such intelligent systems was prone to some problems, creating bottlenecks in terms of widespread availability in classrooms. In order to mitigate this, authoring research emerged. Despite successes recorded in both ITS and authoring research, it was observed that most studies in the field were driven by the intent to test AI techniques and cognitive science principles to the detriment of educational goals, although most of this testing was undertaken in the educational context.

However, current research interest considered the formalisation of the design of ITS/Authoring tools via educational theories, acknowledging the interplay between theory and practice. Consideration was given to the conception of a pedagogic metamodel underpinned by pedagogy theories, which is open to implementation in

ILABS. ILABS represents a platform to test the assumptions of the metamodel and other research issues that were identified. The following chapter, therefore, discusses the theoretical frameworks upon which this research was undertaken, and also the conception of a pedagogic metamodel from the synthesis of the commonalities and differences of the theoretical frameworks, as well as its implementation.

Chapter 3: A Pedagogic Metamodel—Conception & Implementation

In this chapter, an extensive discussion of this research's theoretical frameworks was undertaken (see sections 3.1 - 3.3). As a key research contribution, section 3.4 presented the conception of a pedagogic metamodel, including the examination of its philosophical assumptions; equally, the theoretical elements that constituted the pedagogic metamodel were identified. The implication of the theoretical elements with respect to the implementation of the metamodel was stated in sections 3.5. In order to test the implementation of the pedagogic metamodel, the following were undertaken: an intelligent learning activity builder system (ILABS) and an Intelligent Learning Activity Management System (ILAMS) were developed; thereafter, intelligent learning activity tools (ILATs)—covering marginal costing topic and aimed at learners—were constructed through ILABS by authors. Thus, ILABS—the practical implementation of the pedagogic metamodel—constituted the ITS authoring tool used by authors (i.e. lecturers) for constructing ILATs and was discussed in section 3.6. Also, section 3.6 provided insight into the construction of ILATs. On the other hand, ILAMS—usable by authors and learners—constituted the launchpad for ILATs constructed (via ILABS) for both desktop and/or online learning, as well as managed the inventory of ILATs. The ILAMS was discussed in section 3.7. Finally, the discussion in this chapter addressed one of the objectives of the research, stating how theory shaped ITS authoring design or practice.

3.1 Learning Theories—Basis for A Pedagogic Metamodel

In chapter two, the need for theory-based formalisation of ITS/Authoring tools design was discussed. This necessitates the conceptualisation of a pedagogic metamodel that can inform such a design. Two theoretical frameworks, CT (Pask, 1975) and CA theory (Collins, Brown & Newman, 1989), were identified to underpin the conception of the metamodel. They were chosen on the basis of their philosophical assumptions, which align with the current research context, and would enable investigation of the research issues addressed. Moreover, Pask derived CT on the basis of problem solving and learning (Boyd, 2004), which constitutes the context of this research. On the other hand, CA provides the framework that was scaled up to investigate cognitive visibility (i.e.

making learner's thought processes open) in an environment involving learning by activity (i.e. problem solving) to abstraction, as depicted in Collins, Brown & Holum (1991, p.3) "b*y bringing these tacit processes into the open, students can observe, enact, and practise them with help from the teacher and from other students.*", then useful feedback could be given to enhance learning.

Also, their choice is informed by previous studies that examined and implemented these theories in varying research contexts. For instance, the effectiveness of conversation strategy, for achieving learning (or "coming to know"—Scott, 2001b; Laurillard, 2002), have been investigated in contexts that involve learning via: mobile devices (Sharples, Corlett & Westmancott, 2002; Sharples, 2002; Sharples, 2005); CAL (Scott, 2000; Hartley, 2010); ITS (Patel, Kinshuk & Russell, 2000; Patel, Scott & Kinshuk, 2001); and blended learning (Heinze, Procter & Scott, 2007; Heinze & Heinze, 2009). Also, CA was examined in contexts, such as in: case-based learning (Wang & Bonk, 2001); plan recognition (Connati & VanLehn, 1996); and Web-based/Online Learning (Liu, 2005; Parscal & Hencmann, 2008).

In the first instance, in the cases mentioned above, these illustrate the various dimensions that the theories employed. Secondly, they indicate the relevance of these theories in the current research, because they aim to investigate similar philosophical assumptions, but in a different context, domain and application. The theories therefore constitute the basis to investigate the theory-based formalisation of ITS/authoring design through a metamodel, and the examination of the learning impact of the model's theoretical constructs in the context/domain of this research. So, the above stated research contexts are further examined below to demonstrate how they inform the current research.

Accordingly, CT informed the development of a conversational framework, known as Laurillard's (2002) conversational framework. This framework has been widely cited in the technology-enhanced learning domain for higher education, especially in the United Kingdom (Mayes & Freitas, 2004). It attempts to illuminate and provide conceptual understanding of CT. However, it has been criticised for lacking wide use in practice (Dyke et al., 2007 cited in Heinze & Heinze, 2009). This was attributed to a number of reasons, which include lack of practical considerations, deficiencies in accommodating assessment, and increasing the need to integrate face-to-face sessions into its e-learning implementation (Heinze, Procter & Scott, 2007). Also, Brewster (2009) cited in Luckin

(2010) describes Laurillard’s conversation framework as “*an idealised abstract representation*”, which is useful for understanding how a software system operates in an ideal situation, but not for planning real world interactions. The implication, therefore, is that attention should be given to how much of reality can, and should be modelled (Luckin, 2010), when developing a theory-based model that will inform the design of useful technology-rich learning applications.

Drawing from the above criticisms and scope of the framework, this work finds the framework unsuitable. This is because the conceptualisation of CT in this research aims to address tutorial sessions that may exclude the involvement of human tutors, which is not catered for in Laurillard’s conversation framework. Likewise, the framework does not provide the conceptual basis to investigate the learning impact of cognitive visibility—a key issue being examined in this research. Also, this research intends to enable the practical implementation of a student-system learning interaction, capturing as much real world learning reality as possible, for which Laurillard’s framework is deemed not suitable (see Brewster, 2009 as cited in Luckin, 2010). Therefore, this calls for a new brand of metamodel, underpinned by learning theories that overcome these shortcomings. In order to achieve this, the research focuses on the theoretical assumptions and some theoretical elements of the theories under consideration, which are deemed practically feasible to implement, taking into consideration Luckin’s (2010) argument stated above.

Apart from the above, CT has been used in combination with other theories (or in other contexts) to pursue certain research objective(s) (Boyd, 2004). Heinze, Procter & Scott (2007) investigated the suitability of Laurillard’s Conversation framework, while searching for a theory that could underpin a part-time information technology course in a blended learning context. They examined CT-related literature; their action research data revealed theoretical alignment with blended learning, but showed the weakness of the framework with respect to the research context. This necessitated an amendment and enrichment of the framework, which led to the development of another framework based on CT. However their framework was also not suitable for this research, because it does not provide for some issues being investigated, such as the impact of cognitive visibility on learning. Boyd (2004) claimed that Vazquez-Abad & LaRose (1983) developed and researched an operational tutoring system based on CT combined with structural learning theory. It was implemented on the PLATO system to carry out research on the instruction of rule-based procedures in science education. These examples, therefore, indicate the possibilities of combining CT with any other theory/context in a research process.

Patel, Scott & Kinshuk (2001) describe a research product, Byzantium—an Intelligent Tutoring System, which employed the conversational tutorial approach for teaching concepts and skills in accounting. This work is based on Pask's philosophy that "*teaching is a control of learning*", which provides insights into how teaching and learning can be interpreted and modelled as processes of control and communication. This is incorporated in a tutoring system as a "conversation" strategy to actualise pedagogy. They also claimed to have embedded CA features, such as scaffolding and fading, in the implemented system. Also, in Patel, Kinshuk & Russell (2000), the same CA-based learning-system conversation for cognitive skill acquisition was discussed. Both studies demonstrated the utilisation of some CT and CA features in ITS, employing a model-tracing methodology (i.e. a goal-oriented approach for tracing the solution path).

In contrast to the above, this research utilised both theories to inform the conception of a metamodel. This metamodel offers authors (i.e. lecturers), rich and optionally-reusable theoretical elements that are embeddable in an ITS. This proposed metamodel then forms the basis for the development of ILABS that generates ITSs, employing three optional tutoring strategies—no tutoring, model-tracing and process monitoring (i.e. the process of making a thinking process visible, known as cognitive visibility). Through the latter strategy, this research investigated the feasibility of making a learner's cognitive process visible, thereby confirming the effectiveness of the metamodel as a whole, or at least, some of its features.

Also informally, some studies had deployed the pedagogy strategy of CT, "conversation" or "dialogue", in their work (see Graesser et al., 2001; Graesser et al., 2005; Roque & Traum, 2007; Johnson, 2010). Johnson (ibid), for instance, employed AI methodology to implement conversation strategy in a game-based learning and intelligent tutoring context without any formal reference to CT. This was undertaken to assist learners acquire basic communicative skills in foreign languages and cultures. Experimental results revealed the positive impact of the conversational game-based strategy on motivational effects and learning outcomes. This indicates that conversation has the potential to enhance learning, which forms one of the bases for its consideration in this research. However, this research contrasts with the latter's (i.e. Johnson, 2010) conversational game-based approach, in the sense that it intends to construct ITSs that support the conversation-cognitive tutoring approach. This would enable investigation of the impact of cognitive visibility on learning.

Graesser et al. (2005) examined an AutoTutor system that adopted a conversational strategy to simulate a human tutor by holding a conversation with the learner in natural language. The system employs an animated conversational agent and three-dimensional (3-D) interaction simulations, to enhance the learner's engagement and the depth of learning. They claimed that the work was grounded and motivated by constructivist learning theories with reference to explanation-based constructivist theories of learning, without explicitly stating the theories that supported such a pedagogical strategy. Findings show that this system, employing conversational strategy, helped students to construct knowledge actively (Graesser et al., 2001). Although, this work further demonstrates the significant role of conversation in learning (using natural language), the current research's implementation is conceived as a non-verbal conversation with cognitive visibility capabilities (for the reasons stated earlier). Furthermore, the success stories of the implementation of CT or its pedagogical strategy, either in formal or informal theoretical contexts, and on an individual basis or in integration with another theory, gives credence to its inclusion in the metamodel discussed in this research.

On the other hand, CA theory has been used as the theoretical foundation of many studies (see Edmondson, 2007; Roque & Traum, 2007). In some cases, it has been combined with other theories (see Dennen, 2004). There had been different levels of implementation, with some studies adopting a holistic approach to the educational application of CA process; some studies investigated portions of the process—such as scaffolding or mentoring, while some examined the theory's activities within the community of practice (Dennen & Burner, 2008). In Liu (2005), CA was the main and only theoretical basis of a study to investigate the impact of Web-based learning on pre-service teachers' performance and attitudes towards instructional planning. CA informs the development of a Web-based learning model that integrates expert teachers and Internet technologies. In order to test the model's effectiveness, a Web-based course was designed and a field experiment conducted. Findings show that the course based on a cognitive model more effectively improves pre-service teachers' performance and attitudes on instructional planning, than a traditional course. While the latter study considered CA in a general e-learning environment, this research examines CA within ITS authoring research.

Similarly, Dickey (2008) examined the integration of a model, mainly underpinned by CA theory, in a Web-based course. The research adopted a qualitative methodological framework using an interpretive case study. It aimed to draw out the impact of CA methods on student learning processes of (i) technology skill; and (ii) technology integration methods of teaching. Findings revealed that students were positively disposed to modelling, coaching, scaffolding and exploration. The methods help foster skills, knowledge, and understanding of integrating technology for teaching and learning. However, the last two models do not fit into the context of the current research; neither did they accommodate other theoretical assumptions (e.g. conversation) that underpin it; so, they could not be considered.

Contrasting with the last two studies, Wang & Bonk (2001) combined CA with case-based learning to design a framework to examine the effectiveness of a Groupware-Based Learning Environment (GBLE). They argued that any such design must be underpinned by learning theories to substantiate their effectiveness. Based on this, the principles of CA and case-based learning were integrated to develop a framework. This was implemented in form of a system titled "*Using Notes for a Case-based Learning Environment*" (UNCLE) to attest the framework's utility. An empirical study carried out showed that the theoretical base of CA provided coherent guidance to practice. It also opens up opportunities to fine-tune the pedagogy of case-based learning. While the model supports the socio-constructivist context of learning—an element of the metamodel conceptualised in this thesis, it lacks a learner-system context of learning that is provided in the current research. The GBLE model lacks a meta-nature, a provision incorporated in this research, and has the advantage of providing rich theoretical elements that can be optionally implemented and investigated. Moreover, it does not provide for conversation in learning, a key element of the current investigation.

Despite the limitations of the above reviewed models/studies, they reflected the possibilities of abstracting a metamodel underpinned by two theories. So, the synthesis of CT and CA in this research is not an exemption to stated possibilities. This was considered on the basis of the fundamental tenets these theories posit, which could enhance the investigation of the issues examined in this thesis. So, the synthesis of these theories, leading to the conception of a metamodel, uniquely identifies with, or provides novelty to this research. With the exception of Patel, Kinshuk & Russell (2000), who claimed to have implemented some features of both theories in their Byzantium ITT, no other study could be traced to have integrated both theories as undertaken in this research. Unlike Patel, Kinshuk & Russell (2000), this research considered the synthesis of both theories to inform a metamodel which was implemented in ILABS that can generate intelligent tutors. The implementation provided process monitoring features, the augmentation of conversation within CA framework, which is not provided in the latter work; thus enabled the investigation of improved cognitive visibility.

Therefore, to demonstrate the conception of the metamodel, an overview of both CT and CA is undertaken in the next two sections. The underlying assumptions, building blocks, methods, and other elements or components of these theories were identified, to enable synthesis into a new model.

3.2 Conversation Theory (CT)

As introduced in section 2.9 (chapter 2), CT is interpreted as a scientific theory of learning and teaching (Scott, 2008; Scott & Cong, 2010), conceived by by Gordon Pask (Pask, 1975) and further developed by the latter, Scott and others (Scott & Cong, 2008). The theory provides a framework to explain learning in both living organisms (human) and machines (computers) (Scott & Cong, 2009; Scott & Cong, 2010). Likewise, CT is regarded as a theory of theory building and can serve as a unifying framework for a wide range of different learning theories, including theories of creativity (behaviourist, cognitivist, constructivist) (Scott, 2008). Due to its cybernetic origin, CT is considered a cybernetic theory of observers (i.e. participants in a conversation) and the communications between them (Scott, 2007). It is also regarded as theory of observers because it explains the observer to himself (Scott, 2008). CT is premised on the idea that reliable knowledge exists, is brought forth, and advances in action-grounded conversations (Boyd, 2004). The theory propounds a radical constructivist epistemology (Scott, 2001b), with a profound constructivist and dialogical approach to knowledge advancement. Base on this stance, Scott (2001b) argued that "having knowledge" is a process of knowing and coming to know, not the "storage" of "representations". Notwithstanding, it is "..*still useful to construct external representations of knowledge and to distinguish between different kinds of knowledge* (Scott, 2001b, p.347).

3.2.1 An Overview of CT

In line with its scientific foundation, CT:

- offers to elucidate how interactions among cognitive systems lead to "construction of knowledge", "knowing" or "coming to know" (a term used by Scott, 2001b; Laurillard, 2002);
- desires to preserve both the dynamic/kinetic quality of conversation; and
- emphasises the need for a "knower" (P-individual) within the communication space (Pask, 1975).
- Recognises the existence of different types of knowledge and proposes an entailment and task structure—a knowledge representation scheme—as a way of analysising and representing different knowledge (Scott & Cong, 2008).
- Proposes 'teachback' as a strategy that demonstrates or enhances effective learning, takes learning from operational knowledge to comprehension level (Scott, 2007).

Expounding further, the theory depicts and describes the emergence of knowledge by way of multilevel agreement-oriented conversations. These conversations take place between participants' processes (P-individuals), embodied in M-individuals (brains, processors and augmentations) (Scott, 2008). This theoretical conception in CT was demonstrated by means of a modelling facility—such as CASTE—that supports suitable communication and learning activity-based environment.

As earlier mentioned in section 2.3 (chap. 2), CASTE is a prototypical system developed to support conversation learning, which represents an embodiment of CT in an artefact (Scott, 2007; Scott, 2008). Pask's motivation for developing this system was to devise a "vehicle for driving through knowledge" (Scott, 2007, p.15). As such, the tutorial system demonstrated the conversation between participants involved in a teaching-learning process. Learners were provided with a description of subject matter as an entailment structure, a structure that showed how discrete topics were related one to another, logically and analogically; each topic had an associated task structure, composed of a set of operations that could be carried out (Scott, 2008). When a learner approached the modelling facility to commence learning, it made available a set of lesson materials, based on the task structures. CASTE implements tutorial heuristics—a key constraint that a learner may not engage to learn a specific topic until he/she demonstrates understanding of subordinate topics Such understanding can be demonstrated by showing how those topics were derived and by way of model construction which instantiate those topics. Thus, with the tutorial heuristic, CASTE monitors the current level of understanding of the subject matter in order to inform which subsets of topics that could be worked on at any instance (Scott, 2008). Experimental studies conducted with the CASTE tutorial heuristics were shown to enhance long-term retention (Scott, 2007).

Pask also observed that participants have different learning styles—“holistic” and “serialistic” learning—with respect to their intention to learn a particular topic and/or subordinate topics within a target domain of conversation. According to the account given in Scott (2008), ‘holists’ take on a global look; they identify the overall structure before committing to learn particular rules by holding several in mind, comparing and contrasting, and formulating and testing complex hypotheses. On the other hand, ‘serialists’ identify and learn specific rules as they progress; thus, their understanding of the subject matter is built step-by-step. To accommodate the needs of holists in CASTE, a map of of the subject matter was provided and permitted students to explore different topics concurrently. For a detailed discussion of the CASTE system, see Scott (2000, 2007, 2008).

The above referred participants, in relation to the theory, could be best understood by considering the following statement:

> *CT asserts that what it is we are mainly helping educate and self-construct is not simply one person but rather a wide variety of interwoven competitive P-individuals, some of whom execute in distributed fashion across many bodies and machines.* (Boyd, 2004, p.179)

By implication, P-individuals may be on diverse but integrated systems (M-individuals) when more than two participants are involved.

Pask was careful not to make a distinction between people and interactive systems (such as machines with processors). This enables its application in human-human (teacher-student) or man-machine interaction (i.e. computer-based teaching systems) (Sharples, Corlett & Westmancott, 2002), all integrated to enhance communication during learning. When contextualised, at its simplest implementation involving two participants only, it could be implemented as just a system, for example, a student learning through an interactive system. In this scenario, the tutoring system stands as the second participant engaging the student in conversation. Then, to realise “knowing”, the

P-individual starts by going into romance (i.e. familiarises) with the topic addressed, negotiates round it and comes into agreement with other participant(s). This process follows a cycle until knowledge grows. This cycle of knowledge emergence is known as Whitehead's learning cycle (cited in Boyd, 2004), and is illustrated below.

Romance > Definition > Generalisation > ......and so on

Figure 3.1: Whitehead's Learning Cycles
Source: Boyd (2004, p.180).

During conversation, conflicts may emerge, which should be resolved for learning to occur. In CT, conflict(s) resolution refers to "agreement" on a shared concept, detected as "understanding" (Pask, 1988, p.84). Their relationship—i.e. between agreement and understanding—is captured in Pask's words thus:

> *...agreement over an understanding'... It involves individuals, each of whom exteriorize his or her understandings and confirms that the other's entailments reproduce his or her own, previously internal, concepts* (Pask, 1988, p.84*).*

Conversation and agreement between participants leads to common concepts sharing. It evolves new participants, the P-individuals—so called psychological individuals by Pask, understood to be "autopropagative" discursive participant procedures-bundles, running or being executed in one or among two or more M-individuals (Boyd, 2004). This occurs due to newly-acquired knowledge or skills, which happens only when strict conversations take place among them, and learning conflicts are resolved amicably on a common ground.

Furthermore, Pask asserted that conversations happen about a domain within which common concepts(s) are shared among participants. The domain should be broken down into topics, which are further broken down into smaller units or shareable concepts with established relations. The whole components of the domain should be organised—from the domain (the head) to the smaller units or concepts. Initially, Pask referred to the maps and representations of topics as entailment structures (Pask, 1988). These topics should be a communicable, shared, or public concepts rather than personal concepts. Later, Pask and his colleagues distinguished between entailment structures and entailment meshes. According to them, entailment structures consist of topics and connections among them. This reflects how they may be derived or understood from other topics. Whereas an entailment mesh recognises that the entailment of one topic from others is a momentary situation (in contrast, not static nor hierarchical) that occurs during action or explanation.

Therefore, from the above conceptualisation of CT, three fundamental ideas emerged: pedagogy communication approach, knowledge representation (KR) scheme, and learning strategies. These ideas are exposited as follows:

- First, the theory prescribes a pedagogy approach based on the idea that learning takes place through the medium of conversations about subject matter (Scott & Cong, 2009; Scott & Cong, 2010). Conversation, as posited by the theory, is bi-directional, it attempts to make knowledge explicit and can take place at different levels: natural language (via general discussion), object language (for discussing subject matter), and meta-languages (for talking about learning/language). Conversation may not necessarily be verbal; it could be gesture, pictorial, or mediated through a computer interface (Pask, 1988).
- Second, Pask (1988) proposed that the subject matter of any learning process can be represented as entailment structures or meshes. These structures can exist in a variety of different levels depending on the extent of relationships displayed. The KR scheme put forward is seen as a major product of the theory by supporters, and is considered to have an advantage over semantic networks and other less formalised and non-experimentally based representation schemes, in that it provides a fine-grain connection of concepts in a domain/topic.
- Third, equally fundamental to Pask's theory is the idea of the "teachback" learning strategy. According to Scott & Cong (2007), the term "teachback" refers to learners' ability to provide verbal conceptual definitions, explanations and justification, and non-verbal demonstration of learning about "why" and learning about "how" respectively. Learning about "why" refers to comprehension learning (or cognitive, conceptual/declarative knowledge). On the other hand, learning about "how" means operational learning (or procedural, performance knowledge). Both terms are conceived to be complementary components of effective learning. Pask believed that if students' can teach back, then the student could transfer knowledge from one domain to a different domain. Also, teachback has been shown to enhance long-term retention. As such, it suffices as evidence to prove that learning has actually taken place. In

order to achieve "teachback", Pask recognised that students approach learning in different ways. So, he classified students according to their learning styles, either "serialist" or "holist" (Pask, 1976c); nevertheless, students can switch style since the learning environment enables this.

Based on the above ideas, Hartley (2010) elucidated two contributions to knowledge attributed to Pask's conversational theory. These contributions can be appreciated through the CASTE system—an implementation of CT. Firstly, the theory recognises students' individual learning styles. Pask distinguishes between serialist—in the way they transverse the entailment mesh—and the holists—those who took a more global approach in their navigation of the system. In order to account for individual differences, the system controls the bow-wave of exploration depending on the student's range and type of coverage of the entailment mesh. The second contribution, the theory advocates the idea of "teachback" mentioned above—a process of externalising learning (Scott & Cong, 2010). This idea is believed to enhance knowledge transfer to other related or non-related domains.

Illustrating concepts embedded in the theory and flow of conversations, Pask developed what he called the "skeleton of a conversation" (see figure 3.2). The skeleton includes the role of a teacher (the domain expert), learner and teacher engagement in conversation with one another. Just as the model (or skeleton) depicts teacher-learner conversation, Scott (2001b) argued that the model attracts two further interpretations:

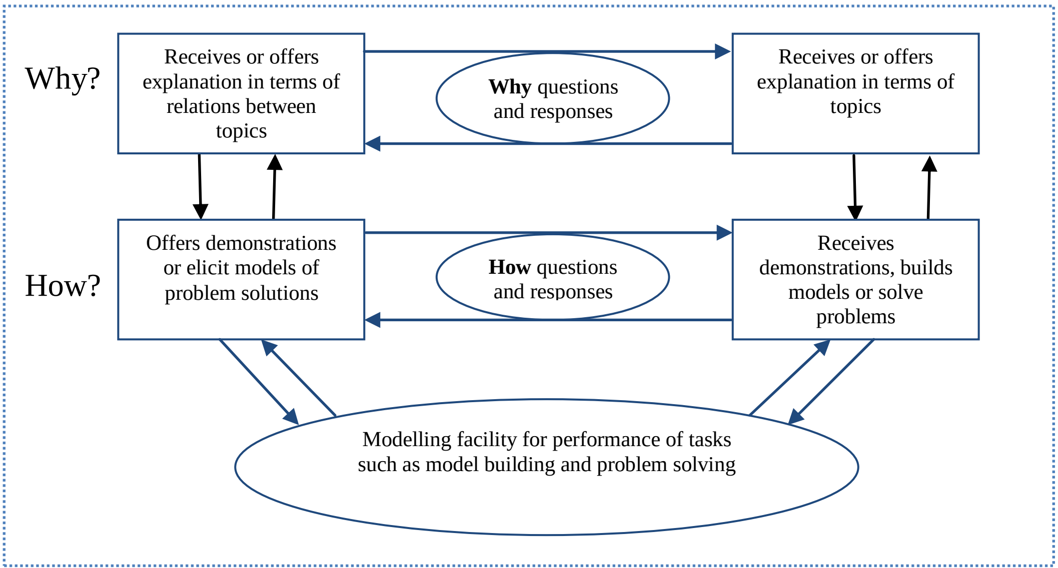


Figure 3.2: Skeleton of conversation after Gordon Pask
Sources: Scott (2001, p.250); and Heinze, Procter & Scott (2007, p.110)

- One, it may be interpreted *"to inform accounts of the genesis of personhood and the 'inner dialogues' that support human learning." (p. 50)*
- Two, *"the model may also be interpreted as showing two peers in conversation exchanging, justifying and demonstrating theories and their associated models and procedures." (p.50*)

The above interpretations provide other outlooks, which were captured in the metamodel developed in this research. They reflect individual development that occurs in the learning space. It depicts an individual learning within his/her internal dialogue, using his sensory organs (specifically cognition) to develop or construct knowledge. This happens through reflection on the subject matter, whereby the individual generates thoughts, analyses them, agrees and disagrees, in order to arrive at a position of understanding. The method, i.e. reflection, is also postulated as a key method of learning in CA theory (that is discussed later in the chapter), could depict conversation as an internal one, involving cognitive parts of individual participants. The other interpretation has some elements that could be classified as collaborative learning—a social context of learning. If this interpretation holds, then it could be said that the theory has some social context implications and should be considered when developing any conversation based tutoring system.

So, in terms of implementation of CT in tutoring systems, it promotes a learning environment that consists of at least two participants (for example, a learner and a teacher), a modelling facility, and at least three levels of interaction (Boyd, 2004). These interactions are: interaction with a shared modelling facility, conversation interaction about how to solve a problem, and why that method should be adopted. The implementation approach is partially assumed in this research in order to limit the complexity associated with the full implementation of the theory due to the time constraint of this work. It suffices for now to show that the theory can be implemented within the context and scope of this research. This type of implementation, therefore, could provide a platform to evaluate the effectiveness of the theory in general, and within the research context under consideration. As such, the following subsections identify the theoretical elements that constitute CT. They are clarified in order to inform the synthesis of CT and CA into a pedagogic metamodel, later in the chapter. Also, their identification would provide means for advancing and rechanneling the application of the theory.

3.2.2 Theoretical Assumptions of CT

As earlier stated, CT is premised on the idea that knowledge exists, evolves and could be advanced through activity-grounded conversation. Based on this, certain assumptions underpin the theory, which were identified by Boyd (2004) and adapted herein. These are outlined thus:

i. the processes leading to the generation of a new mind (or participant) and construction of knowledge can be modelled as multilevel conversations between participants.

ii. The various emergent levels and meta-levels of conversation occur via a communication language that needs to be explicitly recognised, distinguished, and utilised in strategically and tactically optimal ways.

iii. All the constituents of the conversation process—concepts, the memories, the participants and their world-models—can be represented as a collection of procedures (programs) undergoing execution in some combination of biological being (persons) and physical parallel-processing computers called M-individuals.

iv. Useful "strict conversation models" can be made, which bracket off the affective domain, but keep part of the psychomotor and perceptual domain (seen to be an unsatisfactory assumption, but adopted by Pask at the time to enable work to go forward).

v. New P-individuals can emerge when agreements in complex conversations result in a new coherent collection of procedures capable of engaging in further conversations with other such P-individuals.

vi. When such conversation occurs at high enough levels of complexity, it is claimed that a new participant (which can be human actor, team, organisation or society) emerges.

3.2.3 Building Blocks of CT

These assumptions point to twelve constituents—adapted from Boyd (2004)—that should be present in a conversation system/process, thus:

- M-individuals—represent the conversation host(s) or supporting processors, enabling conversation to take place through a network of distributed machines.
- P-individuals—referred to as Psychological individuals by Pask (Pask, 1988); two types of P-individuals can exist: many P-individuals within one person—for

example, a situation of different and possibly conflicting viewpoints running in a single person's brain; and P-individuals made up of many entities—such as schools of thought or organisations when considered from another perspective.

- Formal language—represents the medium of conversation, denoted as L^* language by Pask, should occur at three levels or more in order to avoid what Pask call cognitive 'fixity'—a situation that happens when conversation only take place at two levels.
- Procedures—a set of synchronise-able programs, usually nondeterministic or fuzzy algorithms, which coordinate the learning conversation.
- Stable-concepts—the commonly-shared and public concepts, differentiated from personal concepts, which form the smaller units of the topics learned with the established relationship.
- Topics—represent the essence of conversation (or what is being learned), and encompass all topics in the history of a conversation, captured in the entailment structures or meshes.
- Entailment and entailment structures—capture the derivation of one topic, concepts and their relation with one another; an organisation/order that is not necessarily a hierarchy of prerequisites. It might be conceived as Mind maps (Siemens & Tittenberger, 2009; Buzan & Buzan, 2000 cited in Quinton, 2010) or Concept/Conceptual maps (Pask, 1988; Schmid, DeSimone & McEwen, 2001; Grundspenkis & Strautmane, 2009) of the domain studied. They are not intended to be models of the internal M-individual 'neuro-hormonal' physiological mind that generate processes (Boyd, 2004).
- Task structures—for each topic structure in an entailment mesh, an associated procedure task structure should be constructed giving operational meaning to the topic; the involved task(s) should be uncertainty-reducing.
- Conversation—can be regarded as the teaching and learning strategy, treated as a strict conversation deployed within a fixed agreed domain and conducted as a parallel and synchronous evolving interaction between the P-individuals at a level of language, L_n. It should resolve all forms of uncertainties, such as vagueness and ambiguity (Klir & Weierman, 1999 cited in Boyd, 2004; Pong & Challa, 2007; Kim & Gil, 2008), through questioning and making choices.
- Environment—represents the conversation machine(s) and interfaces that facilitate externalisation of multilevel conversations between/among P-

individuals, similar to some external objects utilised in a traditional learning environment (e.g. paper and pencil, chalk and blackboard—enables all conversations to be externalised).

- Strategies and Protocols—represent strategies deployed to remove conversation uncertainties (termed cognitive 'fixity'—Pask, 1975, p.48; or "person-fixity" - Pask, 1988, p.99); known to hinder further learning progress, a situation that occurs when habits of action and old learning habits (termed 'task-robots' and 'learning robots'—Harri-Augstein & Thomas, 1991 cited in Boyd, 2004) block new learning. Two of these are: fuzziness and ambiguity (as cited in Boyd, 2004). Fuzziness is known to be more troublesome (Brown, Burton & DeKleer, 1982), because it prevents the inflow of new knowledge, since knowledge is seen to be the same all through conversations. So, it should be eliminated through strategies deployed during conversations.

3.3 Cognitive Apprenticeship (CA)

CA theory was developed by Allan Collins, John Seely Brown and Susan E. Newman (Collins, Brown & Newman, 1989). It is regarded as a social constructivist approach to human learning, which advocates a master-apprentice relationship situated in a social context to aid successful learning. It applies an apprentice model (applicable in a psychomotor domain—vocational and trade-based training) to support learning in the cognitive domain. As such, it moves from the world of activity to a world of abstraction or generality. It encourages the more experienced to offer assistance to less experienced people by providing structure and examples to support the attainment of learning goals (Dennen, 2004).

The theory attempts to bring into the open (or externalise) tacit processes (internal in the master and apprentice), so that students can observe, enact and practice them with help from the teacher and other students (Collins, Brown & Newman, 1989). Hence, the theory draws on social interaction that enables the development of cognitive skills through participation in authentic learning experiences (Dennen, 2004), captured in the Collins, Brown & Newman (1989) definition thus: *"learning-through-guided-experience on cognitive and metacognitive, rather than physical, skills and processes." (p.456)* In contrast, although CT encouraged externalisation of conversation and implicitly supports learning collaboration, it does not necessarily imply that all internal processes occurring in individual participants are made visible to other participants, as clearly evidenced in CA theory. This therefore highlights a unique difference in their epistemological/ontological positions, although both hold that knowledge could be advanced through construction.

3.3.1 Overview of CA

As a social constructivist theory of learning, learning emerges from social interactions involving negotiation of content, understanding, and learner needs (Dennen, 2004). It draws inspiration from the traditional apprenticeship (or craftsmanship) approach of Jean Lave (cited in Collins, Brown & Newman, 1989), in which Lave observed how traditional skills (such as tailoring and woodworking) are acquired. In a traditional apprenticeship, the process of carrying out a target skill (i.e. traditional skill) has two features:

- **i.** externality of the learning process, which is available to both student and teacher, enables observation, comment, refinement, and correction; comment—which forms part of the learning process—could be equated to conversation strategy promoted in CT.
- **ii.** A relatively transparent relationship between the process and concrete products that are the outcome of the skill.

Unlike traditional skills, the process of carrying out cognitive skill is hidden (or internal) in both the teacher and student. By implication, teachers have no access to cognitive problem-solving processes taking place in students, making it difficult or impossible to adjust to students' application of skill and knowledge to problems and tasks. Vice versa, students do not have access to the processes of carrying out cognitive problem solving in the teacher, which can form the *".... basis for learning through observation and mimicry.*" (Collins, Brown & Newman, 1989, p.4) In order to address these tendencies, the CA approach tends to *"... bring these tacit processes into the open, where students can observe, enact, and practise them with help from the teacher and from other students."* (Collins, Brown & Newman, 1989, p. 6)

The theory therefore postulates a process in which learning takes place when the master (the domain expert or teacher) teaches the domain skill to an apprentice (the learner). Skills, herein referred to, are conceived cognitive in nature. Applying apprenticeship methods, therefore, requires externalisation of the processes involved in learning. In tandem, the theory emphasises a learning context in which learning is situated, i.e. learning must be relevant to the skills being taught for the approach to be effective and for students to gain mastery of the skills. This is evident and captured in the words of

Brown, Collins & Duguid (1989, p.32), *"Situations might be said to co-produce knowledge through activity. Learning and cognition, it is now possible to argue, are fundamentally situated."* In view of the vital role of learning situations, ignoring it disposes of the purpose of education, i.e. it defeats (kills or neutralises) the goal(s) of providing useable and robust knowledge.

Central to the CA approach, therefore, is 'situatedness' of the learning process and participation in activities yielding knowledge. Situated learning occurs via active participation in an authentic setting, based on the belief that such engagement fosters relevant and transferable learning much more than traditional information-dissemination methods of learning (Dennen, 2004). The emphasis here is on learning that is deeply embedded within an authentic context, which goes beyond learning by doing as promoted in traditional skill learning.

Despite the significance and effectiveness of situated learning as promoted in this theory, the activity that co-produces knowledge is considered a vital component in the learning process. Brown, Collins & Duguid (1989, p.32) expressed its vitality and relationship with learning thus, *"... the activity in which knowledge is developed and deployed, it is now argued, is not separable from or ancillary to learning and cognition. Nor is it neutral. Rather, it is an integral part of what is learned."* Hence, learning activity should be modelled in real-world situations for it to be effective, and teachers should provide a variety of activities that mimic real-world situations for students to practise. They equally argued, *"approaches such as CA that embed learning in activity and make deliberate use of the social and physical context are more in line with the understanding of learning and cognition.."(ibid)*

From the foregoing, both the situational context and activity of learning explain the epistemological position of the theory; they differentiate the theoretical perception of knowing from that of traditional educational practice. In the traditional practice, epistemology *".....concentrated primarily on conceptual representation and made its relation to objects in the world problematic by assuming that, cognitively, representation is prior to all else*" (Collins, Brown & Newman, 1989, p.41). However, in CA theory, the epistemological position holds that *"... activity and perception are importantly and epistemologically prior—at a non-conceptual level—to conceptualisation and that it is on them that more attention needs to be focused." (ibid)* They argued further, an *"... epistemology that begins with activity and perception, which* *are first and foremost embedded in the world, may simply bypass the classical problem of reference—of mediating conceptual representations." (ibid)* Holistically, the theory encourages movement from the world of activity to that of abstraction (or generality), which is illustrated in figure 3.3 below. Based on this epistemological stance, this research envisages a problem-solving context that requires active participation of a learner leading to knowledge construction, varying the complexity and diversity of topic/domain.

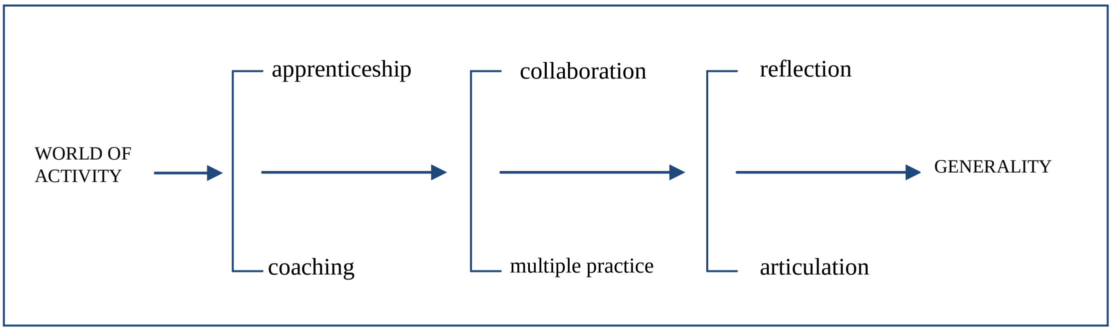


Figure 3.3: Student's Progress from Embedded Activity to Generality
Source: Brown, Collins, & Duguid (1989, p. 40).

3.3.2 Theoretical Assumptions of CA

After a thorough examination of the theory, seven (7) assumptions that seem to underpin the theory could be deduced. These could form the basis for the generation of propositions or hypotheses to test the theory's effectiveness in the context of this research, and in general. The assumptions are outlined as follows:

i. Situation—teaching/learning content (task) must be situated, i.e. it must be relevant to the knowledge and skills required in practice (where it will be deployed);
ii. task—tasks should evolve and be sequenced to reflect the changing demands of learning (by slowly increasing the complexity of tasks so that components skills and model can be integrated—Collins, Brown & Newman, 1989, p.6);
iii. activity & participation—learning evolves through participation in relevant activity, a movement from legitimate peripheral participation (newcomers state) to full or active participation in cognitive activity, conducted in a gradual manner;
iv. context—learning occurs in a social and physical context involving two or more participants (at least one should be an expert assisting less experienced ones, providing structure and examples to support the attainment of goals (Dennen, 2004);

v. externalisation—learning evolves when tacit processes (within the master and the apprentice) are externalised or made visible;

vi. knowledge production—learning situations and activities co-produce knowledge, and are inseparable;

vii. Developmental stages—learning activity and perception are prior to conceptualisation, and should be the focus (encouraging movement from world of activity to abstraction or generality).

3.3.3 Building Blocks of CA

From the above, CA theory could be summed up into four blocks that are required in knowledge construction. They are:

1. Content—the basis or object of teaching and learning. It is composed of knowledge types required to move the novice (apprentice or less experienced people) to expert (master or more experienced people). It includes domain knowledge, and heuristic, control and learning strategies (see table 3.1 below for details).
2. Methodological techniques—touch on the methods adopted to move the knowledge and skills of participants from the actual state (low end) to expected state (high end), leading to the development of both cognitive and metacognitive skills of the novices or apprentices. It consists of six methods: modelling, coaching, scaffolding, articulation, reflection and exploration (further discussed in section 3.3.4 below).
3. Activities sequencing—refers to the ordering of learning activities. The theory posited that learning activities yielding new knowledge and skills should be presented in a sequential or orderly manner so as to have positive impact. In order to achieve that, certain principles are applicable and identified; thus, global before local skills, increasing complexity and increasing diversity of activities (also see table 3.1 for detail).
4. Sociological context—points to the social characteristics or context that constitutes the learning environment. It provides participants (two or more) with the socio-interaction required to evolve and advance knowledge and skills.

The above-identified theoretical blocks and their constituents are in agreement with suggestions in the relevant literature (see Collins, Brown, & Hulum, 1991; Collins, 1991a; Ghefilio, 2001). As a follow-up, Ghefilio (2001) contributed a pictorial framework of the theory, showing the building blocks and constituents of one of the blocks (the method block). This reflected the six methods deduced from the theory. However, this research provided an extended version that reflects the components of all the blocks (see figure 3.4). The aim is to capture all the theoretical elements embodying CA theory; these elements partly informed the constitution of a metamodel developed in this thesis. The extended framework thereby provides a comprehensive view of the theory at a glance; it aligns with the breakdown provided in the latter two references. In addition, table 3.1 below provides a summary of the principles for implementing the blocks in a learning environment. This is an adapted version of Collins, Brown & Holum (1991) and Ghefaili (2003), but expanded and detailed to give deeper meaning. In addition, the implication of each item in the full context of the theory is presented.

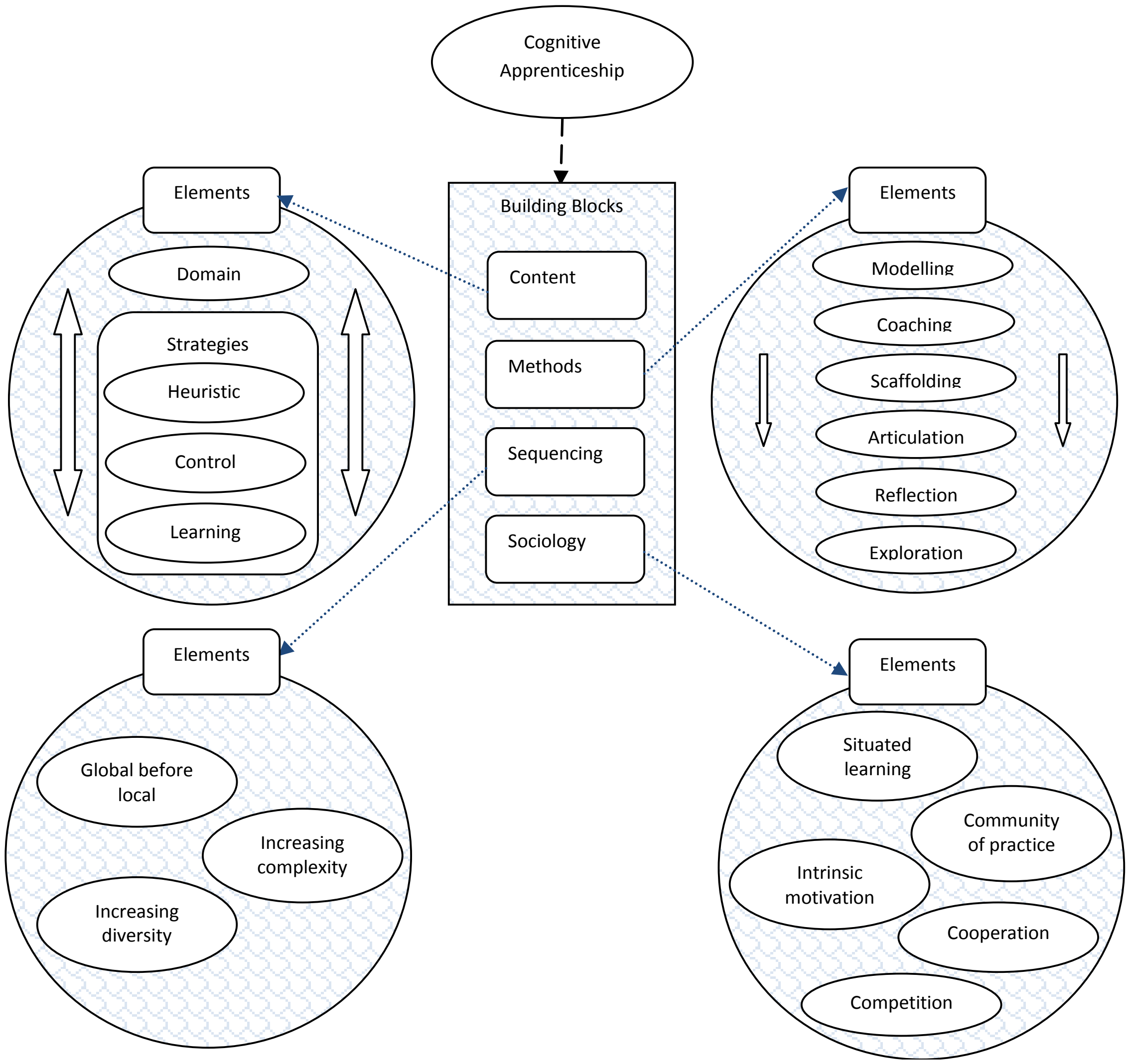


Figure 3.4: An Overview of CA Model (Blocks and Elements)
Source: Extended version of Ghefaili (2003)

Table 3. 1: Principles for Implementing CA Theoretical Elements

CONTENT		**Types of knowledge required for expertise**	**Remark**
1	Domain knowledge	Subject matter, specific concepts, facts and procedures.	• Generally discussed in textbooks, classrooms and demonstrations (Collins, Brown and Hulum, 1991); • Classified as declarative knowledge (from the knowledge dimensions perspectives: declarative and procedural knowledge); • KR scheme will be required in technology-enhanced environment.
2	Heuristic strategies	Generally applicable techniques for accomplishing tasks.	• It refers to what is called "tricks of the trade"; they may not always work, but if they do, they are quite helpful in raising the level of expertise; • It can be developed by experts from the problem solving experience; • There are efforts to explicitly address heuristic learning in the literature (Schoenfeld, 1985 cited in: Collins, Brown & Holum, 1991; Chieu et al., 2010). It has a resemblance with that implemented in the CASTE system, earlier discussed, used to monitor understanding of the subject matter (Scott, 2007).
3	Control strategies	General approaches for directing one's solution process.	
4	Learning strategies	Knowledge about how to learn new concepts, facts, and procedures.	
METHODS		**Techniques to promote the development of expertise**	**Remark**
1	Modelling	Teacher performs a task so students can observe.	• This is the knowledge construction phase in which the

			learner is assumed active
2	Coaching	Teacher observes and facilitates while students perform a task.	• Domain expert is actively monitoring learners' learning process.
3	Scaffolding	Teacher provides supports to help the student perform a task.	• Domain expert provides feedbacks base on learners' learning state.
4	Reflection	Teacher enables students to compare their performance with others.	• Learner reflects on misconception to advance knowledge.
5	Articulation	Teacher encourages students to verbalise their knowledge and thinking.	• The outcome of refection are externalised verbally and/or non-verbally.
6	Exploration	Teacher invites students to pose and solve their own problems.	• Learner attempt new ideas
SEQUENCING		**Ways to ordering learning activities**	**Remarks**
1	Global before local	Focus on conceptualising the whole task before executing the parts.	• This has a resemblance to the holistist and serialist learning styles in CT. They enhance understanding of the target subject matter by enabling the system to adapt to learners based on their learning needs. • While CT considers them as different learning styles and each applied based on learner's style, CA does not. Instead, they are considered as strategies implementable in graduation to achieve understanding irrespective of the learning style of a learner.
2	Increasing complexity	Meaningful tasks gradually increasing in difficulty	
3	Increasing diversity	Practice in a variety of situations to emphasise broad application.	
SOCIOLOGY		**Social characteristics of learning environment.**	**Remarks**

1	Situated learning	Students learn in the context of working on realistic tasks	• Learning is undertaken in context, e.g. through problem solving in applied numerical domains.
2	Community of practice	Communication about different ways to accomplish meaningful tasks.	• Process of exchanging information to advance knowledge.
3	Intrinsic motivation	Students set personal goals to seek skills and solutions.	• This constitutes the inertia to learn that propels the commencement of a learning process.
4	Exploiting cooperation	Students work together to accomplish their goals.	• Synergy effect of collaborative work.
5	Exploiting competition	Positive competition within cooperative learning situations.	• External factor that enhances motivation to learn.

3.3.4 Teaching & Learning Strategies/Methods

CA theory prescribed some teaching methods, which can be used to evolve and advance knowledge and skills. According to Collins et al. (1989), these methods are six in number, and can help students develop cognitive and metacognitive strategies. It also empowers them to use, manage, and discover knowledge. The methods, arranged according to their implementation order, are identified as: modelling, coaching, scaffolding, reflection, articulation, and exploration. Enkenberg (2001) infers that the methods include “explanation”—making seven (7); he asserted that it should stand as a separate (or key) strategy and should come between “modelling” and “coaching” in the order of implementation as listed above. However, Collins (1991a) considered explanation as part or extension of modelling. He claimed that while modelling unfolds the teaching/learning process, explanation provides reasons why it happens. Both views are logical and either of them can be upheld depending on perception and level of implementation details desired. However in this work, the former list is considered concise and precise for implementation. Both methods are considered inseparable in this piece; they work in tangent to achieve the methodological approach posited by the theory. Otherwise, all other methods listed could also be further broken down into sub-methods; for instance, scaffolding was argued to have embedded three critical concepts (see Dennen, 2004, p.815), which can be applied as individual key methods. In order to avoid complexity, since there is no agreed number of methods and to aid the design of a simple and workable framework for this research, the views of the authors of the theory (as stated in Collins et al., 1989; and Collins, 1991a) were upheld. This is taken in the light that there is evidence, and that the same position had been taken in other works (e.g. Ghefaili, 2003).

Below is a brief description of the roles each method played in knowledge construction, with some presented as defined in Enkenberg (2001). The presentation involves some adaptation, by merging modelling and explanation as a method, aligning with the originating view. They are presented thus:

i. Modelling (and explanation)—the demonstration of the temporal processing of thinking, and explaining why activities take place as they do.
ii. Coaching—monitors students’ activities, assisting and supporting them where necessary.

iii. Scaffolding—to support students so that they can cope with the task situation. The strategy also entails the gradual withdrawal of the teacher from the process, when the students can manage on their own.
iv. Reflection—the student assesses and analyses his performance.
v. Articulation—the results of reflection are put into verbal form; although, this may not necessarily be spoken words, but could be gesture and other forms of expression (e.g. non-verbal communication).
vi. Exploration—the students are encouraged to form hypotheses, test them, and to find new ideas and viewpoints (Enkenberg, 2001, p.503).

However, the above list and meanings attached cannot be taken as determinate (or definitive). It is a research effort, which should be acknowledged, moreso, in that there is no standard taxonomy of social constructivist methods in the literature. For example, some refer to mentoring and/or coaching as a form of scaffolding (e.g. McLoughlin, 2002); some consider scaffolding an aspect of coaching (e.g., Collins et al., 1989), while others maintain they are separate strategies (as listed above) falling under larger classification of CA (e.g., Enkenberg, 2001; Jarvela, 1995—cited in Dennen, 2004, p.814). Fading, although not listed above, was also mentioned in the literature and sometimes considered a key strategy/method (e.g., Collins, 1991a; Ghefaili, 2003); or part of scaffolding or one of the three critical concepts―Zone of Proximal Development (ZPD)[1], fading[2], and intersubjectivity[3]―that can be implemented for actualising effective scaffolding (Dennen, 2004). This argument continues in the literature as to the number of methods embodied in the theory, but a definite position should be taken to ease implementation. Therefore, the above six methods were assumed in this work, which expands in detail over the strategies that could be employed to actualise the apprenticeship approach in the cognitive learning domain.

---

[1] ZPD is a dynamic region that is just beyond the learner's present ability level; as learners gain new skills and understanding, their ZPD moves with their development. This space between actual and potential performance is assessed through social interaction between the learner and someone who is more experienced—potentially a teacher, parent, or even an advanced peer (Dennen, 2004, p.215). **[2]**. **Fading** of scaffolding occurs as the learner gains independence and no longer needs support to complete the desired task. (Dennen, 2004, p.216) **[3] Intersubjectivity** – is a shared understanding or goal, lack of which "can be evident in the form of learning conflict, non-participation, or expected outcomes" (Dennen, 2004, p.816). According to Dennen (2004), teachers and learners come to the learning situation with their own understandings and must find a shared meaning to succeed in the learning activity. This shared understanding, called intersubjectivity, is constantly negotiated in our everyday lives, helping in the process of "bridging between the known and the new in communication" (Rogoff, 1990, p. 72 cited in Dennen, 2004, p.216).

Based on the methodological proposition of the theory as identified above, the student's progress can then be represented from situational activity to principles in a sequence that starts with modelling up to exploration. This methodological proposition also supports collaborative learning, as identified in figure 3.4 above. The six methodological techniques are further grouped into three stages in the order of learning progression, with the inclusion of Enkenberg's "explanation" strategy as part of the modelling method. The stages, as identified by Collins et al. (1989), are summarised as follows:

i. Stage 1: Modelling, coaching and scaffolding—are regarded as the core of the apprenticeship approach and help develop cognitive and metacognitive skills. They form the basis of the learning process, and include explanation.
ii. Stage 2: Articulation and reflection—happen to be the next developmental stage of a student as professed by this theory; these help students develop problem solving strategies and execution similar to that of a domain expert (or master).
iii. Stage 3: Exploration—the last method, helps the student develop independence and the ability to identify and solve new problems within the target domain.

3.4 The Conception of an Augmented Conversation and Cognitive Apprenticeship Metamodel (ACCAM)

The above discussions—of CT and CA—centred on their goals, underlying assumptions and their theoretical characteristics, which they propagate to evolve and advance knowledge and skills. Their features, on an individual basis and collectively, introduces theoretical and pedagogical issues believed apt and significant for the conception a metamodel that could stimulate the development of an educational tool.

This section describes the conception of ACCAM—that brings together the characteristics of the theories discussed above, reorganising them and paying attention to the theoretical implications of their synthesised effects. An effort carried out on realising:

- That skills and knowledge taught in schools have become abstracted from their real uses in the world (Collins et al., 1989). However, the current research context/domain—by nature—is practise-based; e.g. accounting, an applied numerical discipline and the research implementation domain, is regarded as a technical and practise-based profession (Jamous & Peloille, 1970 cited in Kinshuk, 1996), so the apprenticeship model of learning involving situated activity seems appropriate.

- That conversation strategy could be useful in evolving and progressing knowledge and skills development, since it has been theoretically and practically asserted that conversation could help construct and reconstruct knowledge, thereby enhancing learning (Graesser et al., 2001; Klemm, 2002; Grasser & D'Mello, 2012).
- The need to develop both the cognitive and meta-cognitive aspects of individual learners, since these elements comes into play during knowledge construction, and have been found effective (du Boulay & Luckin, 2010; Aleven & Koedinger, 2002; Munoz et al., 2006; Roll et al., 2007; Bull & Kay, 2008; du Boulay et al., 2010). These constitute elements of constructivist theories (e.g. CA & CT), which assert that individuals use their personal experiences to make sense of, or seek to understand, the reality that exists (Berger& Luckman, 1966 cited in Darlaston-Jones, 2007).
- The need for a metamodel that could form the foundation of an educational tool that supports the theoretical positions/characteristics of the theories under consideration; since there is interplay between theory and practice (Hartley, 2010), a good theoretical foundation can improve the design of an educational situation (VanLehn, Jones & Chi, 1992).

Equally, current research motivation stems from the need to explore a metamodel that encapsulates and enhances the investigation of issues examined in this thesis, inclusive of cognitive visibility of the learning processes. Realising various uses and applications, to which the theories considered had been subjected in the past—as discussed earlier in the chapter, it is believed that this work could benefit from a metamodel underpinned by these theories. A metamodel constituted by the synthesis of the theoretical assumptions, teaching and learning strategies, and other features proposed by the underlying theories, could reconcile the various positions/elements of the underlying theories to provide common and shareable theoretical elements with agreed interpretation for implementation in an authoring tool. Thus, the metamodel benefits from the individual theories, sharing their common grounds and building on their divergent areas, as illustrated in figure 3.5a below. Figure 3.5b depicts the proposed teaching and learning process of the metamodel. The theoretical elements under consideration, their source(s) and consideration in adjacent theory, and inclusion/exclusion, meaning and implication in the metamodel are presented in table 3.2.

Subsequent sections discuss the characteristics of the conceptualised metamodel. The discussion includes the philosophical stance of the metamodel, a synthesised abstraction of the philosophical positions of the two constructivist theories that underpin it. This suggests an epistemological and ontological position that drives the implementation of the metamodel. Also, other characteristics of the metamodel, drawn from the underlying synthesised theoretical frameworks, are equally discussed. These include learning content, KR scheme, content sequencing, pedagogy methods that include conversation strategy, and the social context of learning. It should be noted that conversation, which is the pedagogy medium propounded in CT, can be internal (i.e. within the P-individual through negotiation of internal processes) or external (i.e. between M-individuals—two cognitive systems). It can also occur in three phases with applicable methods, grouped as follows: phase 1—involves modelling (including explanation), coaching and scaffolding (including fading); phase 2—includes reflection and articulation; and phase 3—involves exploration. In view of the foregoing, the characteristics of the metamodel (as stated in table 3.2 below) are discussed below, stating the standpoint of the metamodel with respect to each.

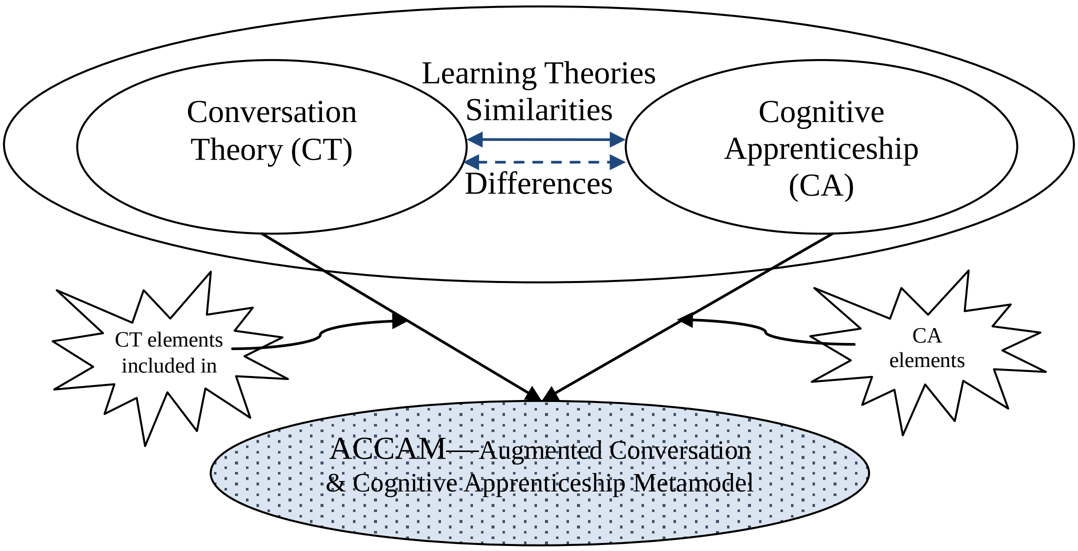


Figure 3.5a: Conception of the ACCAM from CT & CA

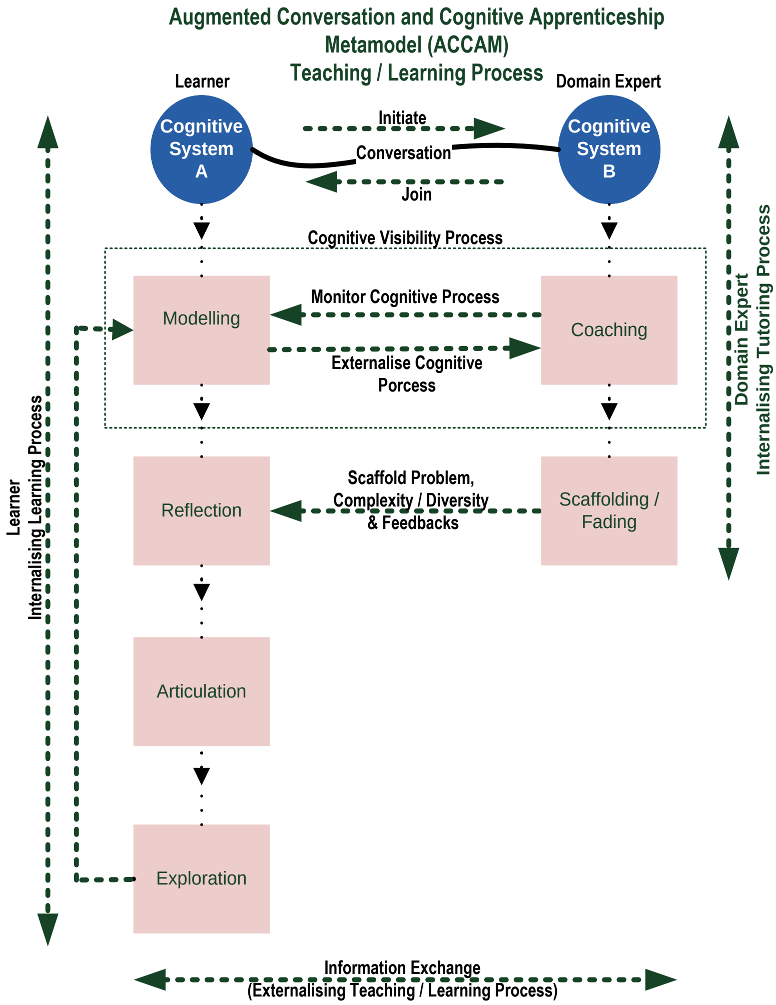


Figure 3.5b: The ACCAM Pedagogy Process

Table 3. 2: Synthesised Theoretical Elements—The Foundation of a Metamodel

Theoretical Elements	Conversation Theory (CT)	Cognitive Apprenticeship (CA)	Augmented Conversation & Cognitive Apprenticeship Metamodel (ACCAM)
Nature of knowledge: The Epistemological/Ontological stance or assumption.	Knowledge exists and evolves, and is advanced through conversation, represented by entailment structures.	Knowledge evolves when situated and advanced through socio-interaction.	Knowledge exists, evolves and advances when situated through socio-interactive conversations with a minimum of two participants. Learning involves establishing relationships among the concepts of the subject matter. Knowledge is structured as a network of interrelated concepts captured as entailment structures.
Content	Conceived as the subject domain upon which learning is based and includes topics and their concepts and task structures.	Conceived as both the subject domain and the three strategies deployed during learning. Domain knowledge must be situated or tailored to practice.	Consists of domain knowledge, captured as entailment structures of topics/concepts, yet situated in practice. Assumes CA strategies, but enhanced to remove conversation uncertainties (termed cognitive ‘fixity’ by Pask).

Theoretical Elements	Conversation Theory (CT)	Cognitive Apprenticeship (CA)	Augmented Conversation & Cognitive Apprenticeship Metamodel (ACCAM)
Strategies and Protocols	Deployed to remove conversation uncertainties, termed cognitive 'fixity' by Pask.	Part of content	Assumes CT, but embedded as part of content.
Entailment and entailment structures	Represent a KR scheme for domain knowledge.	Not prescribed.	Assumes CT definition.
Participants	Conceived as integrated distributed cognitive systems, i.e. M-Individuals and/or P-individual(s), depending on the number of internal processes occurring in the individual's brain during learning. Must include a domain expert and learner(s). Minimum of two participants involved.	Conceived as individuals participating in socio-interaction learning. Must include both experienced person(s) and less experienced person(s) in a master-apprenticeship relationship. Minimum of two participants involved.	Assumed CT meaning with the integrated cognitive systems and/or P-individuals, which embodies the socio-interaction learning context of CA. Must include a domain expert (or master) and learner(s). Minimum of two participants involved.
Procedures	A set of synchronised programs—usually nondeterministic or fuzzy algorithms that coordinate conversation.	Not prescribed.	Included as defined by CT, but implemented as "pure" algorithm (i.e. algorithm that is AI-neutral, instead, interface-based).

Theoretical Elements	Conversation Theory (CT)	Cognitive Apprenticeship (CA)	Augmented Conversation & Cognitive Apprenticeship Metamodel (ACCAM)
Formal language	The medium or language of conversation agreed to by all participants. This should occur in at least three levels to avoid cognitive 'fixity'.	Observation, enactment and practising.	Informal or non-verbal conversation, or what Holland & Childress (2008) regard as information exchange between learner and domain expert, involving active participation (e.g. practising).
Environment	Represents the conversation machine(s) and interfaces that facilitate externalisation of multilevel conversations (i.e. information exchange or bi-directional communication) among participants—similar to pencil and paper, chalk and blackboard etc.	Not prescribed.	Included as defined in CT. Deployed to make visible inner processes taking place in individual participants as required in CA.
Social Context	Learning takes place in an informal social context.	Learning takes place explicitly in a social context.	Assumes both informal and formal social contexts, but limited to two participants (in this implementation).

Theoretical Elements	Conversation Theory (CT)	Cognitive Apprenticeship (CA)	Augmented Conversation & Cognitive Apprenticeship Metamodel (ACCAM)
Teaching and learning Methods	Prescribes 'teachback' strategy to facilitate understanding and knowledge transfer.	Prescribes six methods for moving a learner from actual state (low end) to expected state (high end).	Assumes conversation that involves seven methods of pedagogy (CA methods and CT 'teachback'). It assumes three concepts: Zone of Proximal Development (ZPD), 'intersubjectivity' and fading as part of a scaffolding method. Intersubjectivity refers to having shared understanding or goal, lack of which is evident in learning conflict among participants in a learning situation (Dennen, 2004)
Sequencing	Sequencing of subject matter materials are based on the learning style of a learner. Not just a sequence from global to specific.	Promotes sequencing of learning activity from global to specific.	Integrated both ideas from both theories. Identified the need to recognise the different learning styles of a learner. Despite that, still recognises the need to sequence learning materials in ways that promote understanding irrespective of each learner's style.

3.4.1 The Philosophical Meaning—Ontological / Epistemological Stance

The definition of knowledge and its science of acquisition/construction span the history of teaching and learning theories. These theories provide different dimensions or perspectives to the concept of knowledge, and each suggests different means of achieving the state of knowing (that could be regarded as the essence of learning). These theories range between two extremes of an axis (von Glasersfeld, 2002), depicting the dimensions/views in which knowledge has been perceived, and the often occurring element(s) common to all. One view, the realism stance, holds that what we come to know must be more or less a "true" representation of an independently existing reality; whereas the other end, the subjective idealism, holds that, there is no reality beyond the human mind (von Glasersfeld, 2002). Learning from the history of philosophy, the impossibility of a rationally tenable position anywhere on the established axis of the extremes could be observed, on realising that whatever proposed from one end of the axis has element(s) of the other end of the axis, and could therefore be demolished by grounded arguments (von Glasersfeld, 2002). This informed the need for a specific and clear position (or understanding) with respect to the metamodel under consideration, which is to be implemented in a technology-enhanced environment.

In a technology-enhanced learning environment perspective, three types of knowledge are usually captured: domain knowledge, pedagogy knowledge, and learner's knowledge (Grundspenkis, 2008), but Murray (1998) asserts that pedagogy knowledge includes domain knowledge. The latter claimed ITS should contain expertise on the subject to teach, and expertise on how to teach. The first expertise was referred to as domain knowledge, and the second referred to as teaching knowledge. Ohlsson & Mitrovic (2006) claimed two types of knowledge should be learned, declarative knowledge and procedural knowledge. According to them, procedural knowledge is problem-solving skills and it is differentiated from teaching knowledge. Herein, it is argued that domain knowledge comprises declarative and procedural knowledge, and the latter is differentiated from teaching knowledge, which is conceived as strategies deployed in teaching and learning.

Domain knowledge, classified as declarative and procedural knowledge (Ohlsson & Mitrovic, 2006; Akin, 1986, 2008), could be constructed individually or through external assistance in a socio-interaction as premised in the theories that underpin the conceptualised metamodel. ACCAM, the focused metamodel on which learning would be subjected, assumes three key concepts, namely existence, evolvement, and advancement of knowledge. Therefore, it upheld, epistemologically, views that knowledge exists (aligned to Pask (1989) as discussed above), evolves and progresses within individual participants (Pask, 1989) or through reflection (Collins, Brown & Newman, 1989), as well as through interaction of participants (Pask, 1989; Collins, Brown & Newman, 1989) when knowledge is situated in practice (Collins, Brown & Newman, 1989). According to Pask (1989), this interaction could be a network of cognitive systems, while Collins, Brown & Newman (1989) explicitly based it on sociological interaction. Implicitly, Pask's cognitive systems interaction can be said to be a socio-interaction since participants behind the cognitive systems could be human (the learners) or organisations.

The above stance holds, on the ground that in a numerical problem solving context of a domain that is practice-based or procedurally-oriented, initial romance—the first step in Whitehead's learning cycle (as cited in Boyd, 2004)—with the declarative aspect of knowledge provides information about the target domain, is assumed to take precedence over meaningful procedural learning activity. It brings into existence knowledge that forms the platform for its further evolvement and progression. Hence, the above three concepts could be accomplished in the context under focus, when procedural knowledge is supported with declarative knowledge—assumed prior, and learning is situated to practise and is mediated via conversation in a socio-interaction environment with two or more participants involved.

Declarative knowledge provides an avenue for participants (newcomers, learners or apprentices) to have a romance with the target domain, creating an 'abstract-knowledge-state'. It provides the launch-pad for knowledge development and progression into an 'understanding-state', subject to conditions advanced in the pedagogic metamodel. Existence of the former state, represented as a learner actual state in the ZPD (see figure 3.6 below), does not automatically translate into understanding. This is so, since skills and knowledge taught in schools have become abstracted from their real uses in the world (Collins, Brown & Newman, 1989). It only confirms that the process of knowledge construction has started with classroom teaching, and reflects prior elements of knowledge that could germinate and advance into understanding (the expected state in ZPD), thereby aiding application in practice and knowledge transfer (to diverse topics of the domain). Therefore, bridging the gap between actual state and expected state, as shown in figure 3.6 below, requires situating learning, and encouraging socio-interaction mediated through conversation.

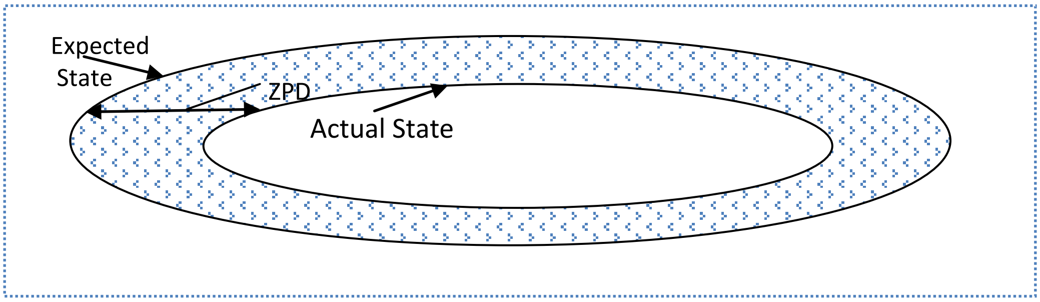


Figure 3.6: Zone of Proximal Development (ZPD)

Relating the above to the current research context, learners are assumed to have received declarative knowledge through the traditional classroom teaching method, thereby providing a basis for existence of prior knowledge. This is then followed by procedural knowledge acquired via the tutoring systems generated from the implemented metamodel. The generated systems provide diverse learning content that is situated to practise, encouraging knowledge transfer. Learning therefore occurs through conversation in a sociological context with a minimum of two participants, in a small-scale implementation of the metamodel. One of the participants represents the domain expert and the other, the actual learner or apprentice. By so doing, knowledge could be evolved and advanced via a technology-enhanced tutoring system that is underpinned by a theory-based metamodel.

Equally, the knowledge in question should be organised or managed in a way that aids construction. Two views associated with management or organisation of knowledge could be identified. Business management scientists understand knowledge management as the systematic process of finding, selecting, organising, distilling and presenting information in a way that improves an employee's comprehension in a specific area of interest. On the other hand, computer scientists conceived it as the organisation of knowledge repositories (databases), to allow for easy retrieval and exchange of the information stored (Li & Masters, 2010). The latter view or definition, which aligns with the proposal of one of the theories underpinning this metamodel, was upheld. It represents the ontological foundation of the metamodel, in which knowledge organisation is understood as entailment structures, as prescribed in CT by Pask (1989), and covering the scope defined in Grundspenkis (2008).

According to Grundspenkis (2008), ontology could take several meanings, which include the following:

- First, ontology could be a knowledge structure; in this case, it not only reflects the domain concepts, but also the relations between them.
- Second, ontology may support reasoning for diagnosis of the causes of the learner's mistakes and misconceptions, seen as a relevant function of the student diagnostic module.
- Third, not only can ontology represent definite concepts and semantics of their relationships, but also all synonyms of both, the concepts and names relationships.
- Fourth, ontology may correspond to taught subjects available on the Internet, usage of which may allow teachers to construct courses reaching compatibility with corresponding ontology, and:
- fifth, each notion of ontology may be supplied with references to corresponding learning objects that may be shown to a student, if the mistakes or misconceptions are detected (Grundspenkis, 2008, p.136).

The above gives a wide coverage of ontology. Although this coverage could be considered in the metamodel discussed herein, it will require enormous resources not available at present. In order to simplify implementation, therefore, taking into consideration the time constraint associated with current research and the contextual aspect, this work adopted ontology as composed of domain topics/concepts and their interrelations. At implementation level, it was represented as a set of interconnected rules used to capture the subject domain, guide diagnosis of the learner's misconceptions, and aid feedback generation.

3.4.2 Learning Content—The Domain Knowledge and Strategies

Domain knowledge constitutes one of the types of knowledge that are captured in a technology-enhanced tutoring system (see Murray, 1998; Grundspenkis, 2008; Woolf, 2009). So, a metamodel meant for such a learning environment should capture domain knowledge as well as the strategies that will be deployed to teach it.

The metamodel under consideration assumed full meaning associated with learning content as conceived in CA. This consisted of the domain knowledge (the focus of teaching and learning) and strategies deployed in the knowledge construction process.

Regarding domain knowledge, it adopted the definition prescribed in CT—a systems theory, since the metamodel is meant to be implemented in a technology-based learning environment, and because CT proposes a KR scheme that can be used to capture knowledge in such an environment, an aspect that is explicitly omitted in CA theory.

3.4.3 The Knowledge Representation (KR) Scheme

Murray (1998) advanced that a technology-enhanced learning environment should capture knowledge of what to teach (domain knowledge) and how to teach (pedagogy knowledge), whereas Grundspenkis (2008) claimed three types of knowledge are captured, namely domain knowledge, pedagogy knowledge and learner's knowledge. However, knowledge has also been conceived to be either declarative (propositional) or procedural knowledge (Akin, 1986, 2008). The latter understanding of knowledge dimensions is argued, not determinate or definitive. For instance, VanLehn (1987) argued from the AI perspective, and claimed that such rigid classification "*is notorious ... as a fuzzy, seldom useful differentiation*" (as cited in Murray, 1998). The latter, therefore, suggested that such classification should be abandoned except in the context in which it has precise meaning.

Whatever position is taken, it is important to note that a KR scheme would be required to store each of the knowledge types. Therefore, to avoid complexity, limit and clarify the knowledge types captured in the KR scheme considered in the proposed metamodel, this thesis adopted the definition of knowledge as comprising domain, pedagogy and learner's knowledge. It assumes domain knowledge to be of two types, i.e. declarative and procedural knowledge. Hence, the proposed metamodel for technology-based learning embraces a KR scheme deemed fit and suitable for the current research context.

The said scheme reflects the structure of the knowledge types within the tutoring system. It shows the knowledge units' inter-relationship, the semantic meaning of the units and their relationship, as well as enhancing access to knowledge during the learning process. In the context of this research, the metamodel adopts the entailment scheme proposed by the CT, although several KRs have been used in ITS research in the past. At implementation level, the entailment structures are captured as a set of rules in which the concepts are interrelated. In order to generalise the metamodel for technology-enhanced learning, the metamodel could adopt any other KR scheme, inasmuch as such a scheme will allow the integration of concepts, and establishes a connection between them.

3.4.4 Sequencing of Domain Knowledge

It has been advanced that human experts did not discover their knowledge or infer it, but learned it from a mentor, either in school or as an apprentice (VanLehn, 1987). Therefore, a good mentor should be careful about selecting tasks that are appropriate for a student's current state of knowledge, because good sequencing of learning content has the potential to maximise or enhance learning (Tedman & Tedman, 2007). Also, VanLehn (1987) argued that learning content should not be in a randomly ordered sequence, but a carefully structured one.

The metamodel put forward in this thesis draws inspiration from the above argument, seen as logically arguable; it holds that students could achieve meaningful learning, when learning content and feedback, hints etc. are structured to their need at every stage of their learning process. So, the pedagogic metamodel assumes a structured sequence that allows knowledge to be presented in a manner that aids its construction. It adopts an integrated sequencing pattern as suggested in table 3.2 above, taking into consideration an earlier-stated suggestion of VanLehn's (1987). Referred pattern allows knowledge to be presented according to complexity and diversity, and tailored to the level and learning style of the learner.

3.4.5 Teaching & Learning Methods

As part of teaching and learning, the methodological approach that a theory prescribes defines how teaching and learning is achieved. Every learning theory or framework, apart from the meaning attached to knowledge, should categorically state how knowledge could be acquired. As such, many theories/frameworks proposed learning methods. For instance, Laurillard's (2002), Scott's (2001) and Heinze, Procter & Scott's (2007) frameworks—underpinned by CT—canvass conversation as medium of coming to know. They upheld the epistemological position of their underlying theory. Thus, the two theories that underpin the pedagogic metamodel discussed in this thesis are not exempted.

In the light of the foregoing, the theories underpinning methodological approaches to learning were taken into consideration to formulate the methods embedded in the metamodel in question. This metamodel hereby marries their learning approaches. It assumes learning could be conducted through conversation (see Pask, 1989), a long-term educational learning mode that has been adopted at several levels of education, including childhood education (see Li & Masters, 2010) to higher education (see

Laurillard, 2002), and found effective. In order to stress its usefulness and relevance in this work, Li & Masters (2010, p.245), states

> *....young children can learn through experience, application, and conversation in community, physically or virtually, with peers, parents, teachers, and other adults, beyond the classroom and across the media.*

Of importance and relevance to this work, is an aspect of the latter excerpt—conversation in a community of practice, which is solidly entrenched in the pedagogic metamodel discussed herein. As mentioned earlier, Laurillard (2002) considered the conversation mode in her framework, which was considered for a higher education level of learning. However, its consideration herein takes a stepwise or phase-wise approach. The steps or phases adopted emerged from the methods proposed by CA theory, a design consideration that is believed could enhance construction of transferable knowledge. Hence, as a matter of significance, relevance and actualisation of deep and transferable knowledge, the metamodel adopts both conversation (CT learning strategy) and the six methods proposed in CA (modelling, coaching, scaffolding/fading, articulation, reflection and exploration) as tools for actualising learning.

3.4.6 Sociology of Learning—The Learning Space and Participants

As earlier quoted above, Li & Masters (2010, p.245) advanced that

> *..... children can learn through..........conversation in community, physically or virtually, with peers, parents, teachers, and other adults, beyond the classroom and across the media.*

This is equally applicable to adults at the higher education level, and has been reflected in several frameworks, some of which were mentioned above and implemented in higher education. Their implementation points to the imperative and relevance of the sociological context of learning, which was considered when conceptualising the current metamodel. In the social context, learning takes place in a space that could consist of learners (peers), parents, teachers and others. Such a learning space should include, at least, a domain expert or experienced person to coordinate the learning process.

The current metamodel builds on the learning environment of the theories discussed and proposes a learning space involving two or more participants engaged in conversation, a learning space, in which teaching and learning is phased in three stages, but directed by an expert (or more experienced people) acting in the space. It aligns with both theories (CT & CA) on the need for at least a domain expert or experienced person, who helps coordinate teaching and learning (see Pask, 1988; Collins, Brown & Newman, 1989). Also, it builds on the traditional one-to-one tutoring environment, in which a more experienced person coordinates or directs a novice or a group of novices in learning a target subject. A pictorial representation of the proposed learning space is provided below (see figure 3.7). It captures how socio-interaction or learning-interaction occurs among the participants in the learning space. However, the extensive implementation of socio-interaction involving more than two participants was not implemented in this research due to its complexity, and because the research is time-constrained.

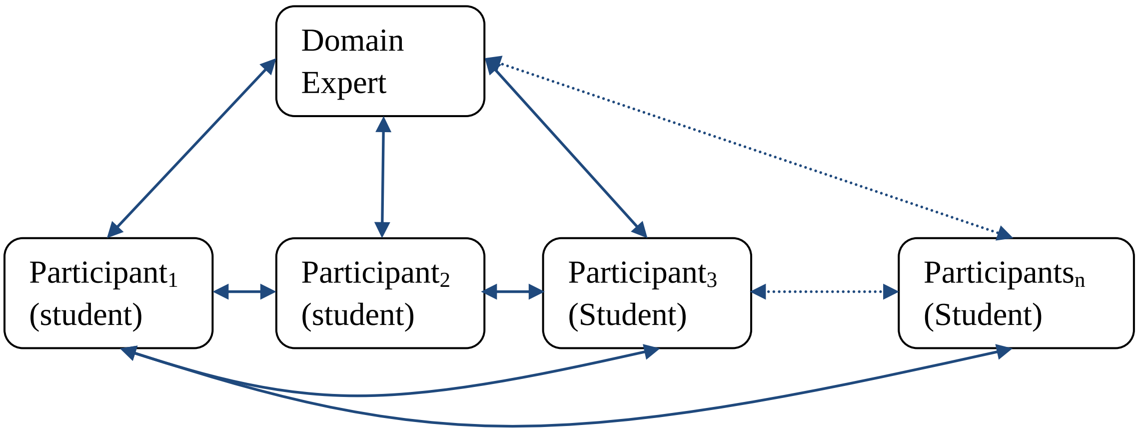


Figure 3.7: The socio-interaction learning space of ACCAM

3.5 Prototyping ACCAM – The Practical Implications

The previous section provided an extensive discussion of the ACCAM, in terms of its ontological and epistemological stance. It also provided insights into elements that constitute the metamodel, which is intended to inform the design of ILABS. In this section, the relationship between the ACCAM and the ILABS is discussed, indicating how the former metamorphoses into the latter. This contrasts with previous work that implemented multiple theories and an ontology engineering-based approach (e.g Hayashi, Bourdeau & Mizoguchi, 2009; Mizoguchi, Hayashi & Bourdeau, 2010) that was considered too complex for non-programmers. Therefore, current work avoids the complexity associated with the latter's approach. It uses only two theoretical frameworks within its ACCAM, transformed into a simple platform (i.e. ILABS) that provides template-oriented selectable features. This platform is utilisable by authors (i.e. lecturers) who are non-programmers, evaluated in the numerical problem-solving context of the accounting and finance domain. It was undertaken in realisation that

Hayashi, Bourdeau & Mizoguchi's (2009) theory-ontological approach has not been evaluated in the context/domain of this research interest, so its suitability cannot be guaranteed in the current research context. Therefore, to demonstrate the current research approach, figure 3.8 below shows the relationship between the ACCAM, and the ILABS and its product (ILAT or ITS).

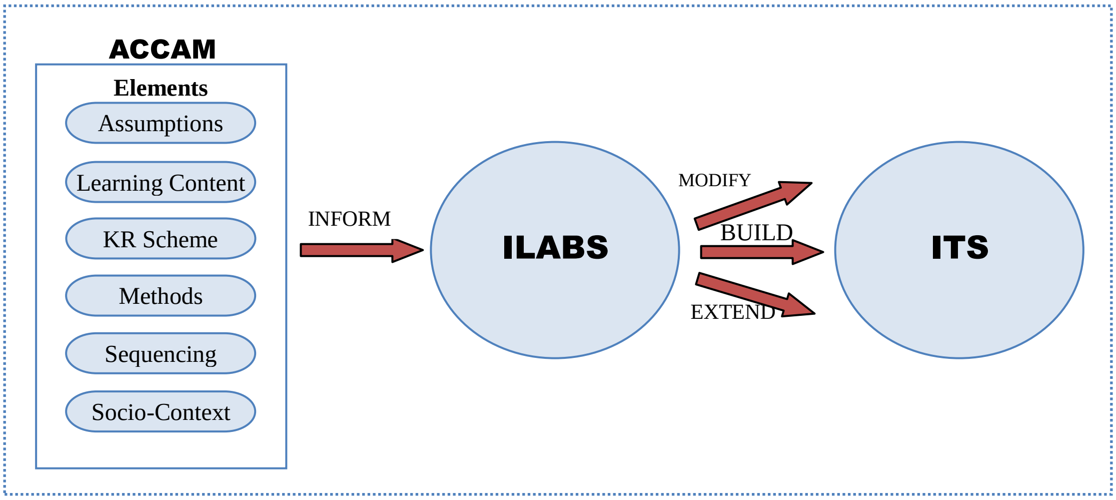


Figure 3.8: The ACCAM and ILABS/ITS Connection

Figure 3.8 thus indicates that ILABS derived its theoretical foundation from ACCAM, a metamodel drawn from two conventional theoretical frameworks—CT (Pask, 1976a) and CA theory (Collins, Brown & Newman, 1989). On the other hand, the ILABS serves three purposes with respect to ILAT construction: build a new ILAT, modify an existing ILAT, or extend an existing ILAT. By so doing, the theoretical elements of ACCAM can be embedded in constructed ILATs through ILABS (the ITS authoring tool and practical implementation of the metamodel). Since ACCAM was designed and implemented in ILABS as a metamodel, authors (i.e. lecturers)—who are non-programmers—can optionally embed any of ILAB's features as best suits their pedagogic goals. This demonstrates the practical implementation of a metamodel in ILABS, undertaken as part of attempts to bridge the gap between theory and practice (see Hayashi, Bourdeau & Mizoguchi, 2009; Hartley, 2010) and in response to Self's (1990b) call for formalisation of ITS design. Thus, the theoretical approach was undertaken purely from an education perspective, in contrast to Self's (1990b) AI perspective, Hartley's (2010) CAI-based theory-practice interplay, and Hayashi, Bourdeau & Mizoguchi's (2009) theory-ontological engineering approach. Although Self (1990b) argued in favour of an AI-driven theoretical foundation, current work adopted an education-based theoretical perspective, since these tools are meant for educational purposes, as argued in chapter 2 of this thesis. It thus establishes the feasibility of a relationship between theory and practice in the ITS/Authoring field from an educational perspective.

As earlier stated, the metamodel is constituted by certain elements, which include: theoretical assumptions, learning content, KR scheme, sequencing strategy, teaching/learning methods driven by conversation, and sociology of learning. The elements were transformed into features captured in ILABS as discussed below.

3.5.1 Implication of the Theoretical Assumptions Implementation

The metamodel assumes that conversation enhances cognitive visibility, which thus enables the generation of relevant feedback in an attempt to enhance learning. This assumption, which consists of two conceptual concepts—conversation and cognitive visibility—informs the inclusion of a calculator feature and three optional tutoring strategies in ILABS, any of which can be optionally embedded in ILAT constructed via the ILABS. Accordingly, ILABS supports the following three optional tutoring strategies: model-tracing, process monitoring, and no tutoring. Model tracing is a goal-oriented knowledge tracing strategy; it compares a learner's solution to a problem-solving goal with that of the domain expert. On the other hand, process monitoring involves comparing each cognitive node or step with that of a domain expert in an attempt to address a problem-solving goal. Hence, the latter strategy is regarded as a step-wise tracking of learning process. So, both strategies compare the learner's solution with a domain expert's version, but at goal and step levels respectively, and then provide appropriate feedback (see VanLehn, 2006—feedback types). However, the "No Tutoring" strategy allows a student to explore a given problem without any guidance during the learning process. Feedback is only provided at the end of the learning process, when completed work is submitted for marking. Therefore, the "No tutoring" strategy assumes a summative assessment, while the other two strategies assume formative assessment—but at differing levels. In that sense, while the "No tutoring" strategy enables the evaluation of the conceptual knowledge/skill a student has acquired over time, the other two strategies—model tracing and process monitoring—evaluate the gradual development of knowledge/skill at two different levels—goal and step levels respectively.

In contrast to previous works that have strong links with cognitive science principles and/or AI techniques (e.g. Aleven et al., 2006c; Zarandi, Khademian & Minaei-Bidgoli, 2012), the conceptualisation and implementation of the tutoring strategies employed in this research derived their theoretical foundation from the education–oriented metamodel, ACCAM. This provided a strong educational basis to evaluate educational research issues, unlike cognitive or AI-based studies that attempt to test cognitive/AI models or techniques within the educational domain. Also, while the model-tracing concept was implemented in the Byzantium project discussed in chapter 2, current work enhanced previous work by augmenting the conversation-based system with the cognitive visibility concept, which is implemented as process monitoring. That said, to enable cognitive visibility of a student's learning process, the process monitoring strategy was strongly tied to the implementation of a calculator. It was through the latter that the cognitive process during numerical problem solving can be monitored. In that regard, the implementation of process monitoring in ILAT requires the embedment of an improvised virtual calculator, unlike the other two strategies (i.e model-tracing, no-tutoring)—in which the calculator is optional. Thus, embedded strategies translate into different tutoring behaviours in the constructed ILAT during learning.

3.5.2 Implication of Learning Content / Knowledge Representation Implementation

Learning content principally represents the knowledge of the subject or topic addressed in a domain of interest. In this research context, it covers only the numerical topics within numerical disciplines. In order to validate the implementation of the pedagogic metamodel, the numerical aspect of the accounting and finance domain was chosen to represent the evaluation domain. This aligns with the trend in the literature, in which various research issues or new approaches were tested in certain domain(s) (e,g. Chi et al., 2011; Gibert et al., 2011; Dewan, 2012). Although, in the near future, the evaluation may be extended to other numerical disciplines involving categorisation and/or application of rules—covered by the implemented metamodel (i.e. ILABS), this will further enhance the generalisation of ILABS. Also, the consideration of a certain scope within numerical disciplines aligns with the arguments of Virvou & Moundridou (2000) and Murray (1999) that the applicability of an ITS authoring tool may not be feasible for all possible domains, and so, should be limited to certain knowledge types. According to the latter, it would enable the production of usable and powerful ITSs.

In relation to the above, domain knowledge forms part of the domain module of an ITS, one of the four-component structures implemented in this research (as discussed in chapter 2). Thus, in line with the entailment structure—the KR scheme of the metamodel, current implementation captures domain knowledge as nodes (or units or components) in a network of interconnected relationships. This is similar to Patel & Kinshuk (1996) —see figure 3.9 below, but implemented in fine grain detail to enhance both goal and step-wise tracing of knowledge, in contrast to only goal-tracing in the latter. So, a unit can be an arithmetic operator or any other operand (e.g. "+", "-", "*", etc.), a variable—regarded as a container that holds a value (e.g. cost, sales, quantity etc.), and a predefined or imported function(s). So, variables, operands, functions and the relationship (i.e. operators) that exist between them are captured as learning content.

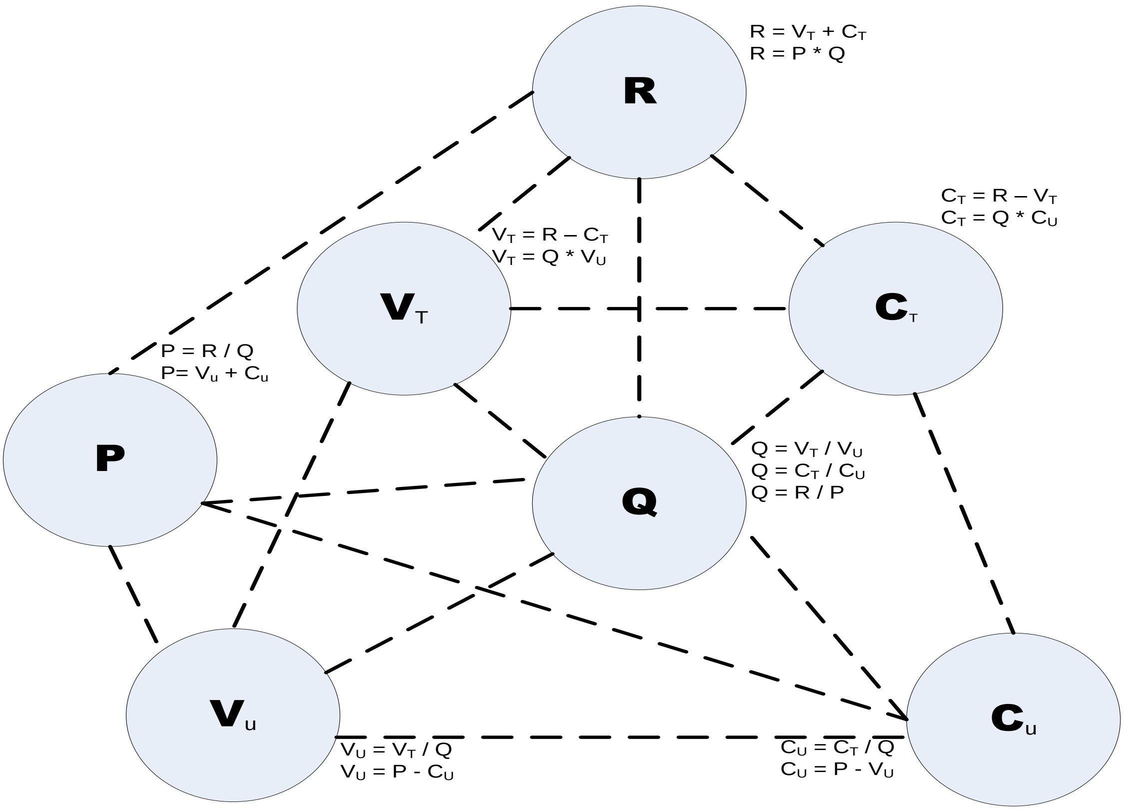


Figure 3.9: Network of interrelated nodes of domain knowledge
Source: Patel & Kinshuk (1997)

3.5.3 Implication of the Sequencing Strategy Implementation

Implementing this aspect of the metamodel, ILABS was developed with template-like features that enable the configuration of problem templates with different levels of complexity and diversity. This enables alignment with the argument of VanLehn (1987) that learning content should not be in a randomly ordered sequence, but a carefully structured one. Accordingly, problem templates are code named, and assigned complexity and diversity codes in an ascending level of complexity—from the least complex to the most difficult problem. However, prior to creating problem templates, problem diversity group names must have been created if more than one diversity group is to be authored; otherwise the system assumes that all problem templates belong to the same group—i.e. are not diverse. Authors (i.e. lecturers) are allowed to specify how the constructed ITS generates problems during learning. Problem complexity and diversity are provided either in a specified structured pattern, sequentially or randomly, but problems within a complexity and diversity group are randomly generated. Each problem is globally presented, with all the variables involved presented to the student on screen. A learner is then expected to perform localised problem solving, reflecting on the relationship between the variables as a whole, then address them, one at a time, without any ordered sequence enforced.

3.5.4 Implication of the Teaching-Learning Methods Implementation

Underlying the ACCAM is the conceptualisation of teaching and learning via conversation and cognitive visibility, identification of misconception and missing conception, scaffolding and fading of guidance etc. Conversation could enhance cognitive visibility—as earlier argued above, and has been considered in varying educational scenarios, from childhood learning (Li & Masters, 2010) to higher education (Laurillard, 2002). Both Laurillard (ibid) and Li & Masters (2010) acknowledged that conversation involves at least two participants, one of which should be well-versed in the target domain.

Also, conversation could be verbal (Graesser et al., 2001; Rudman, Sharples & Baber, 2002; Sharples, 2005; Graesser & D'Mello, 2012) and non-verbal (Scott, 2001b; Scott & Cong, 2010). The current research considered the latter type of conversation, and involves at least two participants—the learner and the tutoring system (containing the domain module that houses the stored expert knowledge). Accordingly, conversation is considered as an information exchange (Klemm, 2002; Holland & Childress, 2008), involving bidirectional communication and interactive exchanges between learner and the tutoring system (in terms of inputs and feedback generated by the system). In order to implement this, a set of generic algorithms were developed, which form part of the underlying elements of the ILABS. These algorithms are embedded in an ITS depending on the tutoring strategies configuration adopted in ILABS during authoring. They are also responsible for the various aspects of the learning, such as cognitive visibility, scaffolding, fading etc. It should be noted that conversation plays a vital role in the implementation of these teaching and learning methods in the ACCAM.

3.6 Implementation of ACCAM: The ILABS, ILAT and ILAMS

ILABS is the practical implementation of the ACCAM. It assumes the theoretical assumptions of the metamodel in a practical sense through implementation of its conceptual constructs as discussed above. ACCAM implementation in ILABS enables the construction of ILAT that utilises non-verbal conversation as a medium of learning, and enhances cognitive visibility of the learner's cognitive process, depending on the tutoring strategy route taken. ILABS was developed for the problem-solving context of applied numerical disciplines, addressing the procedural aspect of knowledge construction only. ILABS development assumes that learners would have learnt the declarative knowledge through traditional classroom teaching-learning mode or via textbook reading. Despite that, ILABS allows authors to capture the declarative knowledge or concepts in ILATs—represented as variables, such that users of constructed ILATs can query any variable to deepen their understanding when required. Based on this assumption, tutoring systems (i.e. ILATs) can be constructed via ILABS that support "learning by doing", generate diverse problems and provide guidance appropriately. This saves enormous resources that lecturers would have expended in providing one-to-one tutorial sessions for students, which may not be feasible in today's educational system in the absence of technology-driven tutoring systems due to the growing student population.

In order to achieve the above, ILABS was developed along with a twin application, known as Intelligent Learning Activity Management System (ILAMS). Both applications are desktop-based and can access a remote repository. The former (i.e. ILABS) is an ITS authoring tool, while the latter (i.e. ILAMS) is a learning management system. ILABS enables authors to construct, build and deploy ILAT onto a remote repository. In addition, it enables modification and extension of any existing ILAT constructed via ILABS. Constructed ILATS can only be implemented within the ILABS during construction or through the ILAMS platform for real learning use (this is further discussed below). Both ILABS and ILAMS were developed using Flex 4 and Action Script 3.0 due to their support for open source development. Figure 3.10 below illustrates the ILABS and ILAMS interaction.

In contrast to Learning Activity Management System (LAMS) was developed by Macquarie University E-learning Centre of Excellence (MELCOE), an authoring tool for creating sequences of learning activities based on content and collaboration (Bower, 2009; Cameron, 2009; Dennis, 2009). ILABS—described in this thesis—focuses on the construction of tutoring systems that address the procedural aspect of knowledge instead of content. Indeed, its twin application—ILAMS—manages the inventory of constructed tutoring systems and their users, enabling learning either on a desktop or online. Future extension intends to include the management of social-interaction between more than two participants. So, the work described in this thesis has the distinction that it provides platforms to construct/execute tutoring systems that enable practice problems, that is, to construct procedural knowledge—which LAMS has no provision for, thereby entrenching previously-learnt abstract or declarative aspect of knowledge.

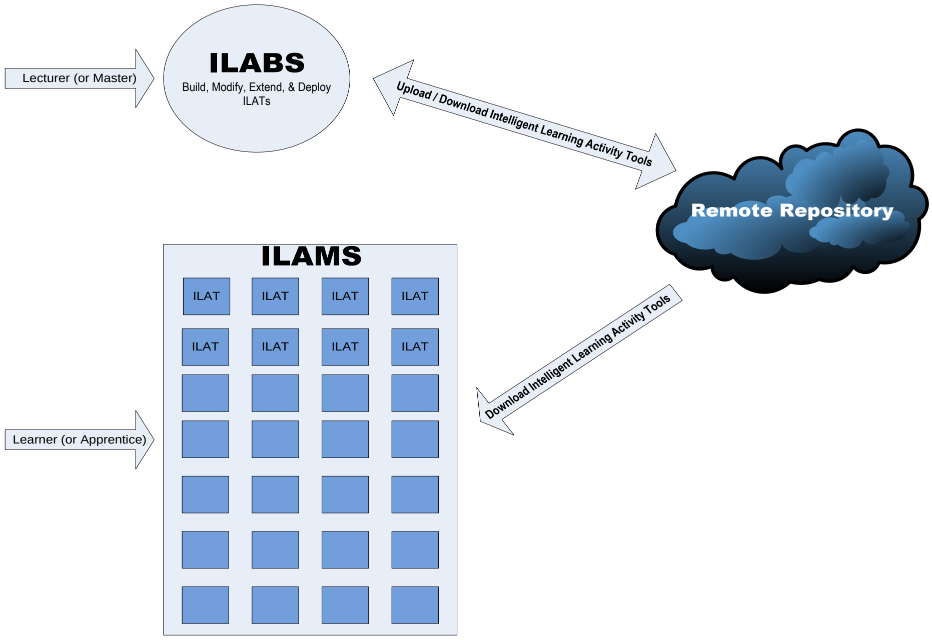


Figure 3.9: Schematic Diagram for ILABS, ILATs and ILAMS Connection

3.6.1 ILABS—The Intelligent Learning Activity Builder System

As mentioned above, ILABS enables authors to build and deploy ILAT (a tutoring system) onto a remote repository, as well as to modify and extend existing tutoring systems. ILABS adopts and implements a four-component structure of an ITS (domain module, student module, tutoring module and interface module—reviewed/discussed in chapter 2 of this thesis).

In accordance with the stated ITS structure, ILABS enables authors to configure the domain knowledge, the user interface and to select one out of the available tutoring strategies to be implemented in an ILAT. The domain module comprises the domain-specific knowledge and problem templates. Domain knowledge represents the core knowledge of the subject or topic addressed in a tutoring system, represented as domain rules and captured in the form of an entailment-like structure (i.e. a network of interconnected components) as proposed in the underlying metamodel. Problem templates enables the generation of practice questions in constructed ILAT, which are meant to deepen understanding of the target domain through provision of unlimited and diversed questions that learners can engage in. In addition, ILABS supports a generic engine that maintains information about a student's knowledge/behaviour during learning, as well as implements-configured tutoring strategy. Figure 3.11 below illustrates the main architecture elements of ILABS.

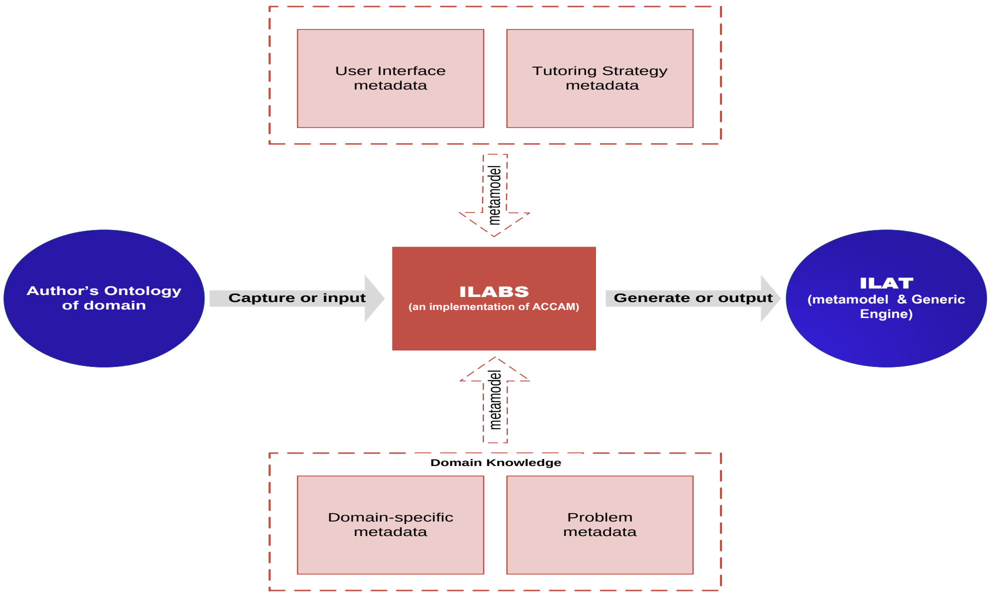


Figure 3.11: System Architecture of an Intelligent Learning Activity Builder System (ILABS)

3.6.1.1 ILABS Features

In line with the above architecture (i.e figure 3.11), the ILABS enables authors to only construct the domain knowledge and the user interface of an ILAT. With respect to tutoring and student modules, three standardised tutoring strategies (informed by the pedagogic metamodel) are supported by ILABS, coupled with a generic student modelling engine (that monitors and stores student learning pattern). The modelling engine is linked to the tutoring strategies. It behaves according to the selected tutoring strategy in order to provide appropriate guidance. Figures 3.21 and 3.22 (in appendix 3.1) represent the screenshots for capturing the domain-specific and problem metadatas respectively, while figure 3.13b—below—shows the tree structure nodes/assets that constitute the interface of a marginal costing ILAT and corresponding view when rendered. Also, figure 3.12 below shows a window with a panel providing the selectable tutoring strategies

ILABS provides menu-driven and template-based interfaces. This includes a new module window, where the initial settings of a new ILAT could be specified, such as the module unique code name, interface type (e.g. single or multiple screens), standardised learning objects and/or functions (e.g. tutor-based calculator, graph functionality etc.), applicable tutoring strategies (e.g. process monitoring, model-tracing and no-tutoring strategies), etc. Equally, it provides a window of drag and drop assets or widgets that could be used to construct the interface. It has a property template, where the properties of assets—utilised in the course of ILAT construction—are specified or modified. Lastly, it provides a window where the domain knowledge can be captured and stored. It also supports a window for specifying problem templates with varying complexity and diversity. Aside from the features presented, ILABS provides other features that enhance authoring of ILATs through its menu-driven design. These include menu options for deployment of ILAT constructed, import of new functions, etc.

Thus, the design of the ILABS has a simple look, which enhances the authoring ability of non-programmers. As a result, programming skill is not required to utilise it. Figure 3.12 below shows the opening window of the ILABS.

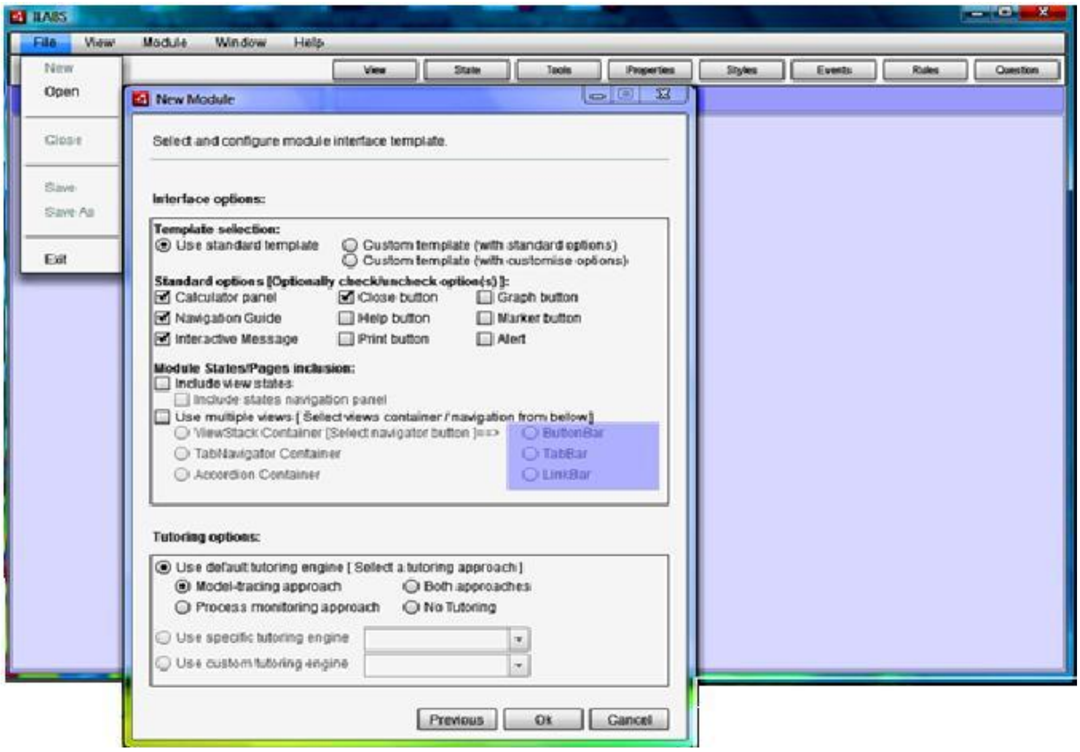


Figure 3.12: Screenshot of the Opening Window of ILABS

3.6.1.2 Using ILABS to Construct an ILAT

In order to author an ILAT, the following processes or steps are undertaken:

i. Launch the ILABS application―click the start button of windows, then select ILABS from the options listed under programs, or double click the ILABS icon on the desktop.
ii. From the "File" menu option, select the "New" option to commence construction of a new ILAT (i.e. a new module).
iii. Assign a unique name to the new module or ILAT and set its parameters based on available options in the window.
iv. Drag and drop assets onto the tree structure on the left hand side of the design interface from the widget window and position each asset according to how it would appear on the ILAT interface.
v. Set the properties and styles of each asset.
vi. Click on the "render" option from the "Module" menu. This renders the visual look of the ILAT interface and enhances judgement in terms of the look and feel of the interface.
vii. Repeat steps [iv] to [vi] until satisfied with the look and feel of the ILAT interface.
viii. Click the "Rules" button on ILABS' window to create domain-specific knowledge (i.e. rules that drives the ILAT―applies to text box assets only).
ix. Click "Question" button on ILABS' window to create problem templates.
x. Test run the ILAT by clicking "Run" in the "Module" menu option of the ILABS window.
xi. Repeat steps [iv] to [x] until satisfied with configured ILAT workings.
xii. Thereafter, build the ILAT and deploy to remote repository using the "Build" and "Deploy" options under the "Module" menu.

The above steps are pictorised in a flowchart format as shown in figure 3.13a. The figure shows typical authoring stages undertaken by an author when constructing an ILAT. Based on the foregoing, this research demonstrates the ILABS authoring process and its capability by creating an ILAT for marginal costing, a topic in management accounting. This was undertaken to illustrate the ability of ILABS to construct/couple together the four components of an ILAT. Figure 3.13b provides a sample screenshot for a marginal costing module (or ILAT) rendered during construction. Appendix 3.1 provides a detailed illustration/discussion of the use of ILABS for the construction of an ILAT by one of the authors involved in this research, while appendix 3.2 demonstrates a typical use of the ILAT by one of the student users.

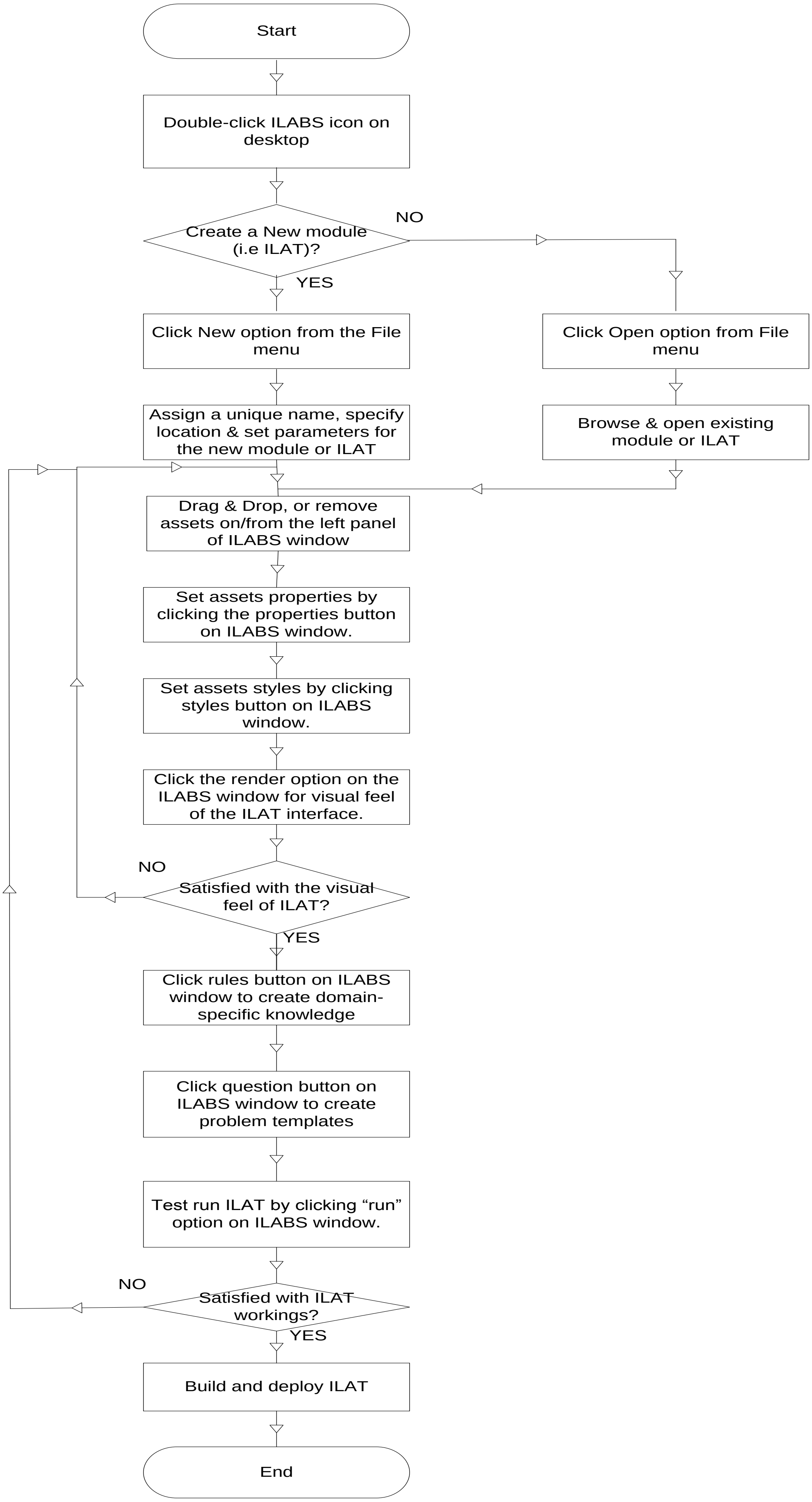


Figure 3.13a: A flowchart of ILAT Authoring Process

For the purpose of this research, the ILAT constructed via ILABS included the implementation of PM which was not considered in existing Byzantium tools. While the Byzantium implemented only model-tracing strategy, constructed ILAT implemented dual-strategy—PM and model-tracing. This enabled the implementation/investigation of improved cognitive visibility through the implementation of PM via a calculator interface, which was not possible in current Byzantium. Thus, current research enhances previous work (i.e. Byzantium) through implementation of PM, and dual-tutoring strategies which enabled switching from one strategy to another depending on the need of users. Also, current research work focused on ITS authoring—underpinned by a pedagogic metamodel—unlike Byzantium that addressed the design of ITTs from scratch. Thus, the approach provided an apriori link between theory and the design of ILABS and the construction of ILATs.

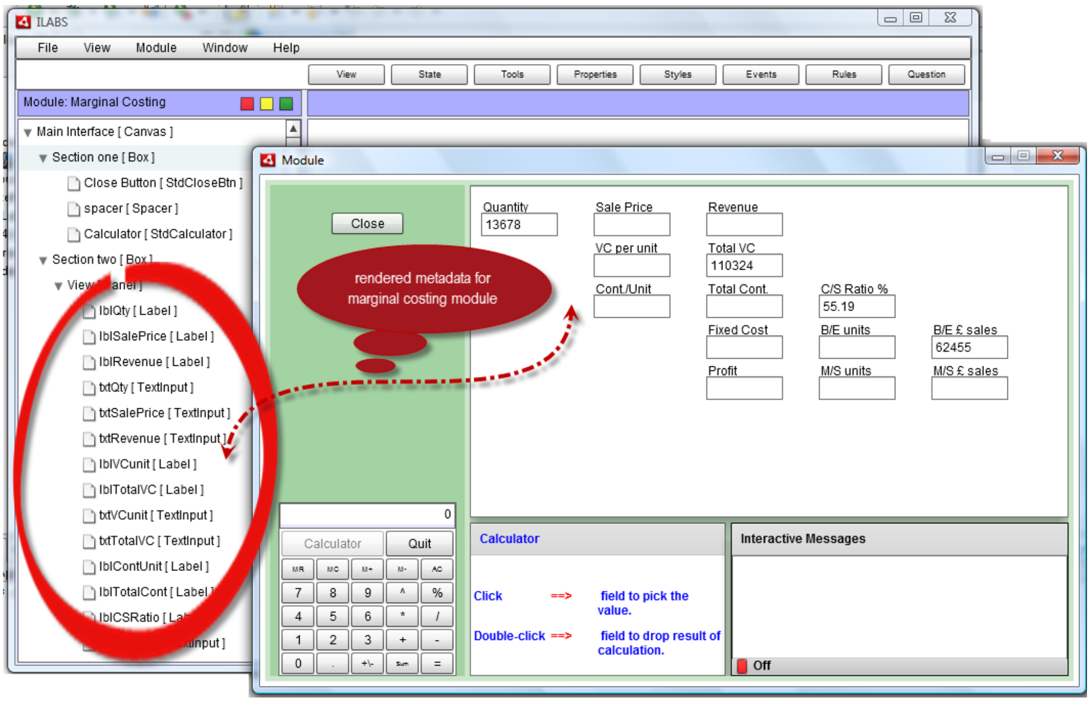


Figure 3.13b: Screenshot of Marginal Costing Module under Construction

3.6.2 ILAMS—The Intelligent Learning Activity Management System

As mentioned above, ILAMS is a learning management system through which users can utilise various constructed ILATSs. Basically, it is expected to perform three main functions: to manage inventory of ILATs, to manage users, and to manage the socio-interaction during collaborative learning involving more than two participants (a function to be incorporated in a future extension of this project).

With respect to the first function, ILAMS provides a platform on which ILATs can be utilised for learning purposes. Accordingly, it supports two learning options, namely offline and online working. In order to work offline, users are expected to access the remote repository via an Internet connection, and then download the required ILATs that can be used via the management system. This option eliminates any bandwidth problem or traffic congestion that may occur during learning, due to a large number of students learning via the Internet. However, provision was made for users, who may prefer online learning through the work online option of ILAMS. Figure 3.14 provides a screenshot of the ILAMS platform with some ILATs already downloaded. So, users need to click an ILAT or the module button to commence learning.

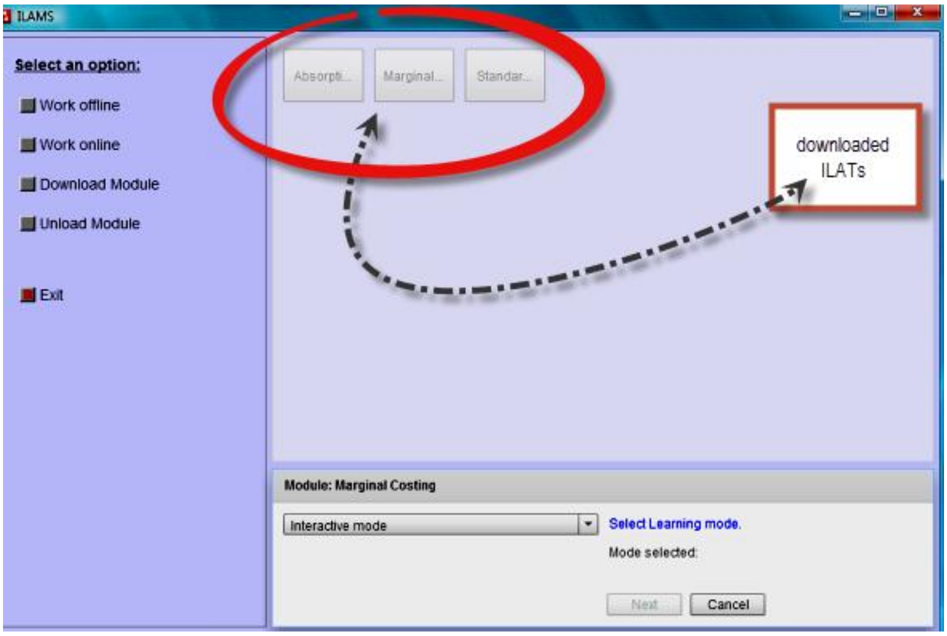


Figure 3.14: Screenshot of Intelligent Learning Activity Management System (ILAMS)

Under the user management function, ILAMS supports four categories of users: administration, lecturers, students and guests. Thus, ILAMS can be used to manage users in terms of setting access rights for various features it supports. Accordingly, each category has specific rights assigned; this determines how users can use the management system and the extent of personalisation that can be undertaken with respect to downloaded ILAT(s). Users with “admin” rights can reset the management system parameters and ILATs, including settings of other user categories. Only the author (lecturer) category has the right over ILATs created by him/herself; thus, a lecturer user can restrict student users, in terms of what can be undertaken in an ILAT or module. Student users, depending on the access rights accorded them, can personalise a ILAT to meet their learning needs. As mentioned earlier, a future extension of ILAMS is envisaged. This would take care of the socio-interaction aspect learning, which is required to implement the socio-learning aspect of ACCAM that provides collaborative learning services. This aspect is considered complex and requires time to implement, which this research could not take on board.

3.7 Summary

In this chapter, an attempt was made to explain one of the fundamentals of the research under consideration by proposing a metamodel of two learning theories. Each of the underpinning theories was extensively discussed, as identified in the literature review chapter. The conception of the ACCAM upholds many of the characteristics of its underlying theories, which were adapted to suit the purpose of this research. Therefore, the interplay of these theories provides a synergistic framework that extensively benefits from its founding theories. Accordingly, the ACCAM was proposed for use in developing an ILABS (or ITS authoring tool in the literature) that could generate an unrestricted number of intelligent tutors in the numerical problem-solving context. This tool, when developed, should provide a practical and easy-to-use authoring environment that captures the characteristics of the metamodel, and enables authoring by teachers or lecturers without programming skills.

Also, the current chapter described the design and development of the prototype ILABS, and its twin application (i.e. ILAMS). ILABS represents an implementation of the metamodel discussed in this chapter, while ILAMS provides the platform to implement the various ILATs constructed using ILABS. Sample ILAT generated from the prototype ILABS were equally discussed. Finally, the following chapter discusses the methodological position taken to evaluate both the ILABS and sample ILAT. Thus, the chapter provides grounds to examine the empirical issues posed in this research with respect to the ILABS and the ILAT.

Chapter 4: Evaluation Methodology

"When the cook tastes the soup, it is formative evaluation; when the dinner guest tastes the soup, it is summative evaluation" (Harvey, J. (ed.), 1998, p. 7)

The previous two chapters explained the research issues related to the field of ITS/Authoring. Also, the theoretical assumptions underpinning ACCAM (a metamodel) were treated. Hence, as a matter of clarification, this research embodies two broad aspects:

- First, the conception and implementation of ACCAM which is meant to explore some research issues that emerged from the literature review in chapter 2.
- Second, the evaluation of the implemented ACCAM (i.e. ILABS—a prototype ITS authoring tool), and the tutoring system (i.e. ILAT) generated from it. This is undertaken to provide explanations for the research issues addressed in this thesis.

Thus, this chapter examines the empirical methodological issues relevant to the second aspect of this research, whereas chapter 3 discussed the conception and implementation phase of ACCAM. The first aspect necessitated an empirical evaluation in order to [i] validate the implemented metamodel, [ii] determine its impact on target users (lecturers and students), and [iii] to determine the extent of alignment with underpinning learning theories (i.e. to either confirm or refute the metamodel assumptions). In order to accomplish the evaluation objectives, the research methodology undertaken with respect to the empirical aspect of the research, and the methods and process employed in collecting and analysing data, are discussed in the following sections.

4.1 Evaluation

ITS and authoring research aims to provide artefacts for educational use. Advances in this research field yielded some tools used in school systems and higher education institutions (Graesser, Conley & Olney, 2012). As a result, attempts to evaluate their reliability for educational use employed different evaluation methodologies (Mark & Greer, 1993), as evidenced in several empirical studies (e.g. Kinshuk, Patel & Russell, 2000; VanLehn, 2011). This research, therefore, is not exempted, since it was undertaken for educational purposes.

However, before planning and implementing any evaluation process, it is necessary to understand what evaluation entails, its scope and purpose, and the tools available to execute it. Also of importance is to determine how an evaluation process should be undertaken, i.e. the steps or procedure needed to actualise the goals of an evaluation. Rindermann (2002) defined evaluation as a description of *"...the systematic analysis and empirical research of activities and programs, their concepts, conditions, processes, and effects.*" (p. 309). Arruabarrena et al. (2002) asserted that evaluation is the process of gathering data, meant to determine the quality or the value of an instructional strategy, and its strength and weakness. Also, Trochim & Donnelly (2008) defined evaluation as *"the systematic assessment of the worth or merit of some object."* (p. 352)

However, the latter definition is considered deficient, since some studies involve descriptive and/or implementation analyses, which may not necessarily relate to worth or merit. Notwithstanding, one common theme from the above definitions is that evaluation aims to gather, process, and interpret information, to aid decision-making. Thus, successful evaluation relates to the object and context in which it is applied. It determines variables or issues that come into play and how the evaluation process contributes to knowledge/practice. However, issues being evaluated may relate to whole educational systems, a component of an educational system, algorithms constituting a system, or abstract and practical issues.

Trochim & Donnelly (2008) claimed that there are different types of evaluation depending on the target object and purpose of the evaluation. Also, several studies claimed evaluation can be approached from different dimensions (Mark & Greer, 1993; Murray, 1993; Draper et al., 1996; Arruabarrena et al., 2002; Rindermann, 2002). However, a number of researchers upheld a formative-summative evaluation classification (Scriven, 1967; Mark & Greer, 1993; Rindermann, 2002; Trochim & Donnelly, 2008; Steiner & Hillemann, 2010). They argued that one or both evaluation approach(es) could be undertaken with respect to a target object (e.g. a tutoring system). The evaluation perspective (i.e. formative-summative) is considered the most important basic distinction often made in evaluation studies (Trochim, 2001; Bennett, 2003).

4.1.1 Overview of Formative/Summative Evaluation

Formative evaluation—sometimes referred to as internal evaluation—is a method for examining the worth of a programme, while the programme activities are forming (or in progress). In the software design context, a system under development is evaluated in order to identify problems. So, information gathered can then be transferred back, to strengthen/improve the system. From the educational perspective, Bennett (2003) claimed that formative evaluation seeks answers to questions about the process of implementation, and how this relates to the achieved curriculum. Steiner & Hillemann (2010) argued that formative evaluation can be used to gather information on the improvement of e-learning technology design aspects during development. In summary, the approach focuses on the building process rather than the final outcome (or product).

In contrast, a summative evaluation is sometimes referred to as external evaluation, e.g. a method for examining the worth of an educational intervention. Usually, it takes place at the end of an educational programme or activities (summation), and is meant to prove or disprove formal claims about the construction, behaviour of, or outcomes related to a completed system. Information gathered through this process provides an overall picture of a finished product, and may be a measure of success (or otherwise) of a product's objectives (Mark & Greer, 1993; Manwaring & Calverley, 1998; Steiner & Hillemann, 2010). According to Saettler (1990), summative evaluation is undertaken to examine the validity of a theory. Bennett (2003) pointed out that it aims to gather data about links between the intended curriculum and the achieved curriculum. Thus, it could be concluded that summative evaluation focuses on project outcome (or final product).

Summative evaluation has a wider coverage, in the sense that all evaluation projects could be subjected to summative evaluation in contrast to formative evaluation. This aligns with Scriven's (1967) argument, in which the latter noted that all evaluations could be summative in nature (i.e. have the potential to serve a summative function), but only some have the additional potential of serving formative functions. Also, Mark & Greer (1993) claimed that because both evaluation approaches are focus-wise different, different methodologies are best suited or deployed. Some addressed internal considerations—such as architecture and behaviour, others focused on external considerations—such as educational impact. Despite the foregoing distinctions, determining the appropriate evaluation methodology to use still poses a challenge. This constitutes a critical issue that must be addressed as part of any research design. On that note, the following section reviews some guidelines that could further enhance the choice of evaluation methodology for this research.

4.1.2 Choosing an Evaluation Methodology

Choosing an evaluation methodology in any given research instance requires clarification of the applicability of evaluation approaches and appropriateness of research methods. In the light of this, Trochim (2001) claimed that the object being evaluated, and the purpose of the evaluation, determines the evaluation type deployed at any given instance. Therefore, to select an evaluation approach, consideration should be given to the characteristics of the approach that best match the purpose of evaluation. An evaluation purpose, usually framed in the form of a research question, drives a research process, which includes the selection of evaluation methodology.

Accordingly, evaluation provides answers to questions it was designed for (Mark & Greer, 1993; Stankov, Glavinic & Grubisic, 2004), and questions asked influence the choice of evaluation methodology (Harvey, 1998). So, a link exists between research question and evaluation methodology (Mark & Greer, 1993; Johnson & Onwuegbuzie, 2004; Silverman, 2010; Trochim & Donnelly, 2008). Accordingly, Trochim & Donnelly (2008, pp.17-18) argued that

> *...what is most fundamental is the research question—research methods should follow research questions in a way that offers the best chance to obtain useful answers.*

Thus, the nature of a research question determines whether a quantitative, qualitative or mixed method/model methodology should be deployed. It informs the use of either formative, summative evaluation, or both. Also, because each methodology represents a different evaluation approach and many approaches are commonly in use, it points to the fact that no single methodology is best (Oliver & Conole, 1998). So, to determine the appropriate methodology to use, research questions should be clarified first, since these could suggest the type of evaluation to employ (e.g. formative, summative, exploratory, experimental etc.).

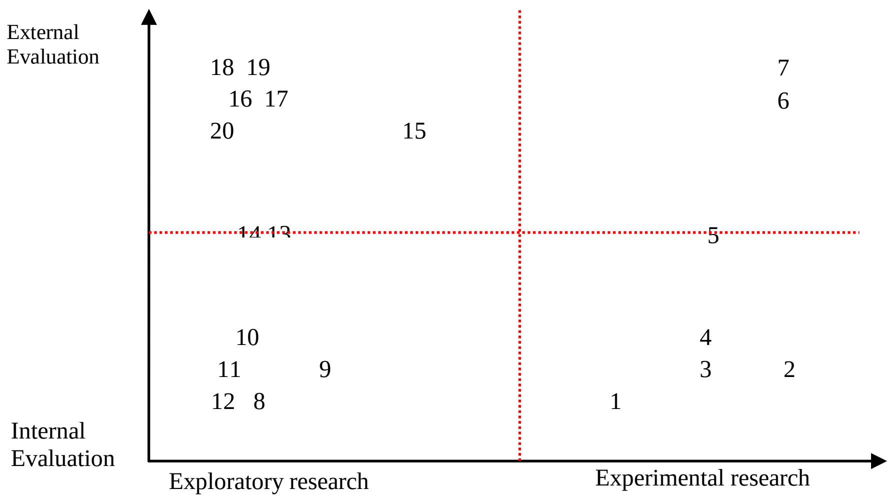
Although a link between research question and methods has been established (Johnson & Onwuegbuzie, 2004), there seem to be no clear guidelines for determining the appropriate method to use in a particular context (Kinshuk, Patel & Russell, 2000; Jeremi, Javanovic & Gasevic, 2009). This suggests the relevance of research context (e.g. ITS research domain) in the research design. Likewise, factors contributing to method selection should include the source and size of data required for the choice to be appropriate in a target context. In the light of the foregoing, Iqbal et al. (1999) propose a classification of existing research methods that could enhance research design, based on two primary questions that every evaluator should answer prior to setting out on an evaluation exercise; these are:

- What is being evaluated: the whole system or just a system component?
- Is it feasible to systematically manipulate variables in the evaluation, and how many users are available for the purpose of evaluation?

Based on the aforementioned questions, existing evaluation methods were classified along two dimensions (Stankov, Glavinic & Grubisic, 2004). Each dimension relates to one of the questions stated above. The first dimension pertains to the degree of evaluation encompassed by an evaluation method. By degree, reference is made to either the whole or part of a whole system. Accordingly, Stankov, Glavinic & Grubisic (2004) and Le & Menzel (2008) claimed that if a method solely concentrates on examining a component or the inner workings of a system, it can be suited for internal evaluation; however, if the evaluation covers the whole system, it is suitable for external evaluation. Both internal and external evaluation could be conceived as formative and summative respectively, depending on the characteristics of the target context in which the evaluation is performed.

The second dimension relates to the feasibility of using a particular evaluation method. It differentiates between a method that attempts to establish cause-effect through controlled investigation (i.e. experimental research), and one that accumulates a large amount of data about a specific aspect of a target object/system (i.e. exploratory research). Experimental research demands the conduct of experiments, and involves the systematic variation of independent variable(s) while measuring the dependent variable(s), ascertaining random assignment of participants to conditions, and requiring statistically significant groups (Iqbal et al., 1999; Ross & Morrison, 2004; Ruxton & Colegrave, 2006; Beaumont, 2009). On the other hand, exploratory research includes in-depth study of the system in a natural context using multiple sources of data. This is usually used where the sample size is small, and the research phenomenon area is poorly understood (Creswell & Plano Clark, 2007).

Thus, the dimensional classifications (see figure 4.1 below) provide four unique groups of methods. Each evaluation method falls in one group or the other. Few methods have attributes of more than a group; such method(s) satisfy(ies) both groups' condition and could be used in either group's context. For example, a method in the borderline of exploratory-internal and exploratory-external evaluation groups could be utilised for both internal and external evaluation in an exploratory research; the same rule applies to the method that falls on the borderline of experimental-internal and experimental-external evaluation groups. This presupposes that each method falls in at least one evaluation group. Therefore, to select an evaluation method, its group methodological requirements must be considered and met. These requirements are summarised in the form of guidelines in table 4.1 below, and can be used to screen methods in any particular evaluation context.



Key to Evaluation methods:

1. Proof of correctness
2. Additive experimental
3. Diagnostic accuracy
4. Feedback/instruction quality
5. Sensitivity Analysis
6. Experimental research
7. Product evaluation
8. Expert knowledge
9. Level of Agreement
10. Wizard of Oz experiment
11. Performance metrics
12. Internal evaluation
13. Criterion-based
14. Pilot testing
15. Certification
16. Outside assessment
17. Existence proofs
18. Observation & qualitative classification
19. Structured task & quantitative
20. Comparison studies classification

Figure 4.1: Classification chart of evaluation methods
Source: adopted from Iqbal et al. (1999)

The above discussion and guidelines (in table 4.1 below), therefore, provide a general evaluation perspective. So, in terms of contextual application, the following section reviews evaluation in the ITS/Authoring research field where this research is situated. In essence, the above and the following sections would enable the grounding of this research within an evaluation framework that best suits it, and demonstrates how this research's methodology evolved.

Table 4.1: Guidelines for Selecting Evaluation Approach

Evaluation Classification	Guidelines
Internal and/or Formative	• Evaluation meant to test individual components, algorithms, technique, approaches, concepts etc. built into, or implemented in, a system; • Evaluation conducted at the beginning and during the developmental stage of the system; • Evaluation results are meant for system improvement, not for claims; • Evaluation focuses on the process rather than final outcome or product; • Not all evaluations are formative or internal;
External and/or Summative	• Evaluation targets the whole system • Evaluation is conducted at the end of the system development; • Design and development claim(s) are formal, and evaluation results either confirm or refute it; • Evaluation focuses on final outcome or product; • All evaluations can be summative in line with Scriven's (1967) argument (even if the evaluation is testing individual system units, yet meant to address overall system goals). • Evaluation meant to test the validity of a theory embedded in the design of an ITS/Authoring tool; • Evaluation meant to determine the impact of an educational practice or paradigm;
Experimental research	• Evaluation involves establishing causality from controlled investigation; • Evaluation is conducted in experimental format; • Evaluation involves systematic manipulation of independent variable(s), while measuring dependent variable(s); • Participants are randomly assigned to evaluation conditions; • Participants are categorised into two or more groups and each group is statistically significant
Exploratory research	• It involves in-depth study of the system in a natural context; • It involves multiple sources of data • Sample size is usually small • Research area is poorly understood

4.2 Evaluation in ITS/Authoring Literature

Evaluation of educational tools seems unavoidable, if these tools are to be deployed for real classroom use. This becomes imperative, considering tremendous research efforts that have been demonstrated in the field of AI in education in the last four decades (e.g. Corbett & Anderson, 1989; Sykes, 2005; Gilbert et al., 2011; Piech et al., 2012; Zarandi, Khademian & Minaei-Bidgoli, 2012). Consequently, several intelligent tutors and ITS authoring tools have been developed by researchers and are increasingly employed in education (Ainsworth & Grimshaw, 2004; Ainsworth & Fleming, 2006; Ritter et al., 2007). These works aim to support learning activities through provision of one-to-one tutoring systems. In some cases, these are aimed at investigating research issues (Mark & Greer, 1993; Muldner & Conati, 2010).

Being educational tools (i.e. ITS and authoring tools), one of the goals of their developers is to build systems that are as reliable and effective as human tutors (Graesser et al., 2005; Dede, 2008; Smith & Sherwood, 1976 cited in VanLehn, 2011). This therefore raises some open questions. How do we ascertain that a developer's goal(s) are achieved? What will inform the reliability of these tools? How do we determine their educational effectiveness, and behaviour with respect to their design requirements? Clues to these questions could be deduced from Mark & Greer (1993). They noted, as these systems are built to investigate research issues, evaluation methodology becomes imminent. This gives credence to the significant role evaluation could play in the success of ITS research. Evaluation therefore provides a platform to confirm design goals, and determine reliability and the effectiveness of tools. It also helps examine their usability.

The foregoing therefore suggest the need to incorporate an evaluation phase(s) in any educational-oriented adaptive systems project, such as the development of intelligent tutors, as well as their authoring tools. Although evaluation phases may address diverse objectives, they should include evaluation issues, such as the effectiveness and usability of the educational intervention (i.e. tools). Thereafter, the efficacy of the system's components in achieving the overall effectiveness of the tutoring system should be examined in the real world of usage (Kinshuk, 1996). In addition, Heller (1991) pointed out that instructional software should undergo some formal evaluation before deploying into the classroom or is used for research purposes (cited in Kinshuk, Patel & Russell, 2000). This is a necessary step, in order not to pass on software that hinders educational goals.

Similarly, Ainsworth & Fleming (2006) argued that to create a learning environment in a time-effective manner, requires an authoring tool that should be easy to learn within a short training period. Its interface should meet authors' needs—simple tools and appropriate feedback on consequences of their authoring decisions; authors should be able to reflect their pedagogic beliefs, as well as meet their learners' needs. Therefore, for the learning environment created from the authoring tool to be effective, the latter claimed learners must be able to understand the subject matter, be motivated and reach learning outcomes in a time-effective way. In order to ascertain the success of the learning outcomes, they argued that large-scale experimental evaluation would be required. Also, the learning environment effectiveness would be influenced by its contextual usage. By implication, therefore, the quality of ITSs generated would depend heavily on the ability of the authoring tool. As such, it should be well designed, developed, and evaluated in order to ensure usability, friendliness and effectiveness.

Also, a survey of ITS/Authoring literature shows that both quantitative and qualitative methods have been employed in various evaluation tasks. Most evaluations relating to the overall effectiveness of tutoring systems preferred quantitative methods (Mark & Greer, 1993; Legree et al., 1993 cited in Kinshuk, Patel & Russell, 2000; and Le & Menzel, 2008), while those relating to the internal efficiency of the whole system, as well as, its individual components, favour qualitative methods (Murray, 1993). Wyatt & Spegelhater (1990) cited in Kinshuk et al.(2000) proposed laboratory evaluation as the most suitable method for the initial evaluation stage (this could be regarded as the formative evaluation stage), while field trials were suggested for later stages of the evaluation process (possibly regarded as the summative evaluation stage). These evaluation approaches, stages and objectives were found to have common across evaluation studies in the field.

In line with the foregoing, Ainsworth & Fleming (2006) conducted an experimental evaluation of their authoring tool, REDEEM, to determine its effectiveness, usability and other learning outcomes mentioned earlier. Findings show that it could be used to author learning environments that are effective, and that it exceeded its initial expectations. They equally noted that improvements to its design could further enhance its functionality. Also, the research findings show that on average, a trained author who is familiar with the domain and with teaching, recorded an average authoring time of four hours per hour of instruction - a ratio of 4:1—to create an ITS from imported domain material.

Jeremi, Jovanovi & Gasevic (2009) assessed the effectiveness of the DEPTHS ITS design, the accuracy of the applied student model, and students' subjective experiences with the system (see also Jeremi, Jovanovi & Gasevic, 2012). They adopted Donald Kirkpatrick's model (see Kirkpatrick, 1979—for a full description) for measuring the effectiveness of a training programme. As a result, two main sets of evaluations were conducted: reaction evaluation and learning evaluation. For the reaction evaluation, they assessed system effectiveness by employing a method involving two steps: first, they analysed students' reactions to the training programme using a questionnaire; second, they conducted an experiment with an experimental and two control groups, then compared pre-test and post-test results of the groups. In order to test student model accuracy, they employed a method which involved the comparison of the system's assumptions about students' performance level to their results on an external test. For the learning evaluation, they employed what they termed *"non-equivalent comparison-group design."* According to Marczyk, DeMatteo & Festinger (2005), this design type happened to be one of the most commonly used quasi-experimental designs for determining system effectiveness. In order to fulfil the experimental design requirements, they ensured an evenly distributed number of students per group. Also, all the students were tested at the outset of the study (pre-test), similar to what was done under reaction evaluation. This evaluation demonstrates the use of multiple methods—involving the use of questionnaire, experiment, pre- and post-test—in a layered evaluation within ITS/Authoring research context, unlike Ainsworth & Fleming (2006) in which experimentation and pre- and post-test evaluations only were utilised.

Weibelzahl (2003) proposed an evaluation framework for adaptive tutoring systems. The framework comprised four evaluation segments: input data, inference mechanism, adaptation decisions, and interaction evaluation. The framework was tested via HTML-tutor, an adaptive tutoring system. The tutor's student model accuracy was assessed using two methods (or steps), by comparing the system's underlying assumptions vis-à-vis (i) an external test, and (ii) actual displayed behaviour of the student. The tutor's inference mechanism was also evaluated. The evaluation carried out attempted to draw a link between the system's assumptions accuracy and the systems' inference mechanism. It did not include or reflect any result on the system's overall effectiveness; instead, it was more of an internal evaluation.

In contrast to the latter, Miller & Butz (2004) carried out external evaluation of a system. Specifically, it was an evaluation of the usability and effectiveness of an Interactive Multimedia Intelligent System (IMITS), a system designed to teach a second-year electrical engineering module. Evaluation of this system was underpinned by two views: (i) the extent of its usability; and (ii) its effectiveness. They used a questionnaire and system log files to collect data. Information on the students' reaction to the IMITS was collected via the former instrument and usability data through the latter. They attempted to determine the impact of IMITS on student learning using quasi-experimental design, similar to DEPTHS evaluation discussed above, but utilised only one control group and one experimental group. Findings revealed that the IMITS improved performance. Also, the students' learning pattern was subjected to regression analysis. This shows that more students used IMITS to learn some engineering concepts. The term "usage," as applied in the research, was defined as a percentage of IMITS's questions presented in relation to a specific concept. Although the evaluation design was similar to that of DEPTHS and their findings were positive, and both evaluated usability and effectiveness of their tutoring systems, an obvious difference observed was the non-inclusion of a comparison of students' pre- and post-test results within the experimental group.

The above highlights some evaluation dimensions, justifications and findings that were traced to ITS/Authoring literature. Also, the link between research questions/context and empirical methods was demonstrated. The above stressed the importance of these connections and the theoretical/methodological assumptions that should form the basis for an evaluation study. Paramythis, Weibelzahl & Masthoff (2010) provided a detailed evaluation framework that reflects the application of empirical methods, quantitative as well as qualitative, showing their points of relevance and the requirements that should be met to utilise them. Aligning the above reviewed works with the research discussed in this thesis, consideration was given to the research objectives driving it. These objectives, which include the conception and implementation of a pedagogic metamodel in an ITS authoring tool, necessitated the validation of the theoretical assumptions of a prototype ILABS and its products (i.e. ILATs). Also, investigation of the perceived learning impact of the metamodel's theoretical constructs was considered, since experiment—the only approach to determine actual impact—was not included in the evaluation process.

The foregoing objectives, therefore, informed two questions addressed in this research, and employed summative evaluation using mixed methods (see Onwuegbuzie & Leech, 2006; and Creswell & Plano Clark, 2007—for types of mixed methods design). Accordingly, the mixed method could involve the use of multiple data sources—such as the use of both quantitative and qualitative techniques—in a single study (e.g. Grant, Kinnersley & Field, 2012). As such, this work adopted the quantitative method on the grounds that it is the most suitable for theory-based research (Chin, Junglas & Roldan, 2012 cited in Conboy, Fitzgerald & Mathiassen, 2012), and could enable the confirmation/refutation of theoretical constructs in a research context—a key element of this research. On the other hand, the qualitative method is noted to enhance the emergence of themes, and deepens insights (Creswell & Plano Clark, 2007; Conboy, Fitzgerald & Mathiassen, 2012), and has the potential of providing answers to the "what" and "how" of a phenomenon (Onwuegbuzie & Leech, 2006), while still enabling its confirmation. In that sense, this research also employed the qualitative method since it sought to understand how the theoretical constructs being examined aid the detection of learners' misconception, feedback generation and learning. Therefore, the use of the latter method could provide reasons/how ACCAM's theoretical constructs (i.e. conversation and cognitive visibility) impacts the above stated constructs, which the quantitative aspect might not capture.

Thus, the use of mixed methods (detailed in Onwuegbuzie & Leech, 2006; Leech & Onwuegbuzie, 2009)—unless otherwise stated—enables the investigation of theoretical constructs in constructivist technology-based learning activities tools, similar to their implementation in a socio-technical Web-based study (e.g. Tinati et al., 2012). So, it enables triangulation of findings from both quantitative and qualitative aspects of this research. As Turner & Turner (2009, p.1) noted, triangulation is "*the means by which an alternate perspective is used to validate, challenge or extend existing findings*". Also, Torrance (2012) argued that triangulation has the advantage of enabling validation of findings from the analyses of data from different sources by comparing the quantitative and qualitative perspectives of users of the tools evaluated. As a result, a better understanding of the issues investigated could be accomplished. More so that both methods have been utilised in varying ways in previous ITS/Authoring studies (e.g. Sykes, 2005; Chi et al., 2011; VanLehn, 2011), although most ITS evaluation studies seem to tend toward quantitative-experimentation. So, in section 4.3 below, the research design/questions are presented and discussed respectively.

4.3 The Research Design

As mentioned above, this research is driven by four objectives, explicitly set out in chapter 1, arising from the research issues and the theoretical frameworks discussed in the chapter 2. In order to address these objectives, two research questions were coined, taking into cognisance some methodological views discussed earlier. For instance, Trochim & Donnelly (2008) noted that the research question, being central to any study, should be framed in the language of the theory that underpins a study, while Silverman (2005, p.77) argued that "*...they point to the methods and data that will be needed.*" Accordingly, the questions being examined were aligned to the research's theoretical frameworks, and phrasing was done to capture all the notable elements of the research's objectives. Furthermore, each question was broken down into four propositions (see chapter 1—section 1.4) to enhance "operationalisation" and examination of theoretical constructs, and to address specific research objectives.

In order to answer the referred questions (see section 1.4), this research takes into consideration the nature of the work undertaken in this thesis—an end-of-product metamodel-based design evaluation. Thus, it employed a summative evaluation—as earlier mentioned—as part of the research design. The appropriateness of summative evaluation was considered from the perspective of Harvey (1998), as quoted earlier. Also, as mentioned above, summative evaluation was adopted in alliance with the research's objectives, which tend to determine if a metamodel can be implemented in an ITS authoring tool, thereby confirming or refuting its feasibility.

On one hand, the foregoing aligns with the argument of Saettler (1990) that summative evaluation could examine the validity of a theory, or determine the impact of an educational practice, so that future efforts may be improved or modified. On the other hand, it could embody formative elements, since some summative evaluations could have the potential of serving formative functions (Scriven, 1967). From the foregoing perspectives, the findings from the adoption of summative evaluation could inform advanced versions of the prototype ILABS in any future investigation, thus serving formative functions in future work. It should be noted that the prototype being evaluated in this thesis is an implementation of the metamodel developed in chapter three, which is meant to generate tutoring systems in the numerical problem-solving context of accounting and finance. Consequent to the adoption of summative evaluation, the research employs a mixed methodology (see Onwuegbuzie & Leech, 2006; Leech & Onwuegbuzie, 2009) as argued above. This involves the utilisation of quantitative and qualitative methods, unless otherwise stated. This methodological position is further explained below.

4.3.1 The Methodology of the Research

In accordance with the summative evaluation adopted, current research involves the collection of primary data. This requires using appropriate evaluation methods/approaches that would enhance the investigation of the research questions being examined. Mertler & Charles (2005) noted that when evaluation is undertaken, it utilises either quantitative, qualitative or both approaches in the collection/analysis of data. In that respect and as earlier argued, the research employed both quantitative and qualitative approaches to gather its primary data. This was undertaken with respect to the propositions treated in this research, unless otherwise stated. The nature of the research and the application of both approaches in current research context (i.e. ITS/Authoring research in the numerical problem-solving domain) demanded that respondents should have hands-on experience with evaluated objects (i.e. the prototype ILABS and tutoring systems generated from it). Thereafter, respondents' perceptions with regard to the evaluated objects could then be captured through quantitative and qualitative instruments. This involved the utilisation of methods/techniques comprising questionnaires and interviews (details on the instruments are presented later in the chapter). The combination of these methods from two distinct paradigms, i.e. quantitative and qualitative research paradigms, suggests the application of a mixed methodology in this research. This methodological position was deliberately chosen to benefit from the synergy or differing strengths and non-overlapping weaknesses of both approaches (Creswell & Plano Clark, 2007; Tinati et al., 2012; Edwards & Crossley, 2009 cited in Tinati et al., 2012). This is congruent with the benefits associated with mixed methodology reported in relevant literature (see Tashakkori & Teddlie, 1998, 2003; Johnson & Onwuegbuzie, 2004; Johnson, Onwuegbuzie & Turner, 2007; Bazeley, 2004, 2009; Denscombe, 2007, 2010).

As Bazeley (2004, p.3) observed,

> *.....often the purpose of choosing a mixed method design is not made clear by the researcher (Greener et al., 1989), potentially leading to confusion in the design phase of the study.*

Subjecting this research to such confusion implies a repeat of earlier researchers' mistakes. Hence, tapping into mixed methods in this research, provided breadth and depth to the issues investigated. It does so by attempting to maximise the variation in the population of the two types of subjects this research focused on. This is because the research tapped into two different user populations (students and lecturers) of varying sizes. While one has a large population, from which a statistically significant sample could be drawn and is suitable for quantitative research (i.e. students); the other has a comparatively small population (i.e. lecturers). Undertaking only quantitative analysis in the latter instance (i.e. lecturers) may not be sufficiently adequate to reach a meaningful/justified conclusion. A qualitative study therefore provides a deeper data set that could be analysed. Moreover, quantitative research is known to give probabilistic meaning to issues. Therefore, incorporating a qualitative aspect stands to benefit from derivable subjective information, and provides another world view and depth to issues under consideration.

Although application of mixed methodology may not necessarily enable the triangulation of findings from both ends of the methodological axis—i.e. quantitative and qualitative—in all situations, since there are can be other reasons for employing mixed methods in a research (Creswell & Plano Clark, 2007). However, this research, attempted to triangulate findings to enable validation of perspectives across the methodological axis, and at the same time enrich the research findings by deepening insights into why and how certain phenomenon occurred. So, the balance of findings from both the quantitative and qualitative data sets, gathered through questionnaire and interview, with respect to objects evaluated by users (i.e. ILABS and tutoring systems), is considered a viable research endeavour that should be explored, and may enhance the reliability of conclusions reached on issues examined in this thesis. Accordingly, detailed analysis of where and how each method was applied in relation to each question is presented below.

Research Question One—For this aspect of the research, both quantitative and qualitative approaches were employed. The research utilised questionnaire and interview instruments to gather data. The focus population consisted of lecturers in the business schools of higher institutions of learning, because: [i] issues examined relate to lecturers in department(s) related to the research's evaluation domain (i.e. accounting and finance department[s]); and [ii] because they are the target users of the implemented metamodel—the prototype ILABS. Due to their relatively small population, compared to the students' population, extensive statistical analysis might be insufficient to reach a meaningful conclusion. Hence, the four propositions (treated within research question one) were examined through questionnaire and interviews, after some exposure sessions on the prototype ILABS. Thus, the qualitative approach (using interviews) played a secondary role, while the quantitative approach (using questionnaires) was the dominant (or primary) technique for data collection (see Leech & Onwuegbuzie, 2009). So, the questionnaire was used as a primary instrument to gather quantitative information, providing quantitative/probabilistic meaning of users' views on issues addressed; while interviews were used to gather qualitative data, providing an interpretive aspect to the findings, and to deepen the investigation. The combination of these techniques enabled synthesis of findings from questionnaire and interview; it provided a platform to reach a reliable conclusion on the theoretical assumptions underlying the implemented metamodel (i.e. the prototype ILABS), ease of use, and usability.

Research Question Two—In this case, the focus participants were undergraduate students taking introductory modules in accounting and finance disciplines, being the main users of the generated tutoring systems (ILATs). The questionnaire instrument was administered because there was access to a large number of participants, sufficient to attain statistically significant samples (Cohen, 1988; Pallant, 2010; Lowenthal & Leech, 2010). In order to secure a statistically significant sample, being a key factor in any quantitative analysis (Omoteso, 2006; Collins, Onwuegbuzie & Jiao, 2007; Lowenthal & Leech, 2010), the data collection procedure was made slightly flexible to attract enough participants. Evaluation sessions were organised and students were made to join any session that was convenient for them. Participants were administered the questionnaire after the exposure sessions.

4.3.2 Population and Sampling Design—The Sample Scheme/Size

In research involving human subjects, participants sampled should be drawn from an explicitly defined population, by stating its characteristics, in order to enhance the credibility of such research (Sapsford & Jupp, 2006). Also, the characteristics of samples drawn should match that of its supposed population to establish credibility (Oppenheim, 1992; Sapsford & Jupp, 2006). Sampling is unavoidable, especially when it is not feasible to involve an entire population. Otherwise, sampling would be irrelevant, e.g. when the population is so small that the entire population can be covered. Sampling is regarded a key success factor in any study, because it helps establish the quality of inferences from the findings of a study (Collins, Onwuegbuzie & Jiao, 2007). Hence, it should be drawn in a way that ensures the credibility of the study. This can be achieved if the sampling process is appropriately determined, i.e. ensuring its alignment to the research goals, research objectives, research questions and the chosen methodology (Lowenthal & Leech, 2010). Also, credibility can be attained, if the sampling process is explicitly stated to allow future replication by other researchers (Lowenthal & Leech, 2010), as well as explicitly stating how other research factors are handled.

In the literature, sampling has been emphasised much in quantitative research, but historically, not much emphasis is given to it in qualitative research (Onwuegbuzie &

Leech, 2007). This can be attributed to a number of factors, one of which could be the number of respondents required, a relatively lower number when compared to quantitative research. However, it is important to point out that sampling is important in all research (Onwuegbuzie & Collins, 2007; Lowenthal & Leech, 2010), be it quantitative research, qualitative research or a research involving both approaches. It could make or mar the success of a research. Due consideration should, therefore, be given to the sample design, which comprises the sample scheme(s) and sample size(s). According to Collins, Onwuegbuzie, & Jiao (2007), a sample scheme refers to the explicit strategies utilised to select units (e.g. people, group, settings and events), whereas sample size indicates the number of units selected for the study.

Also, different sampling schemes have been proposed in the literature, ranging from probabilistic to non-probabilistic schemes (see Collins, Onwuegbuzie, & Jiao, 2007; Onwuegbuzie & Collins, 2007). Several factors dictate what sample size to select, and this includes the research questions and research design used (Lowenthal & Leech, 2010). With respect to quantitative research, emphasis is placed on a statistically significant sample size (Cohen, 1992; Omoteso, 2006; Lowenthal & Leech, 2010); whereas in qualitative research, the guiding principle should be the concept of saturation (Ziebland & McPherson, 2006; Onwuegbuzie & Collins, 2007; Teddlie & Yu, 2007; Mason, 2010). The term saturation refers to the point when you have heard a range of ideas and are not getting new information (Teddlie & Yu, 2007) or "*the collection of new data does not shed any further light on the issue under investigation*" (Mason, 2010, p.2). As such, sample size should not be so small as to make it difficult to accomplish data saturation, theoretical saturation, or information redundancy (Onwuegbuzie & Collins, 2007; Ali & Yusof, 2011). That is, the sample size should be large enough to eliminate subjectivity (Mason, 2010). However, several studies suggested different minimum sample sizes and this constitutes a basic guideline. For instance, a causal-comparative study should have 51 participants per group for a one-tailed and 64 for a two-tailed hypothesis; a correlation study requires 64 for a one-tailed and 82 for a two tailed hypothesis; experimental study should have 21 participants per group for a one-tailed hypothesis (Onwuegbuzie et al., 2004); phenomenological design should have between six and ten interviews (Morse, 1994; Creswell, 1998), etc. —as presented in Onwuegbuzie & Collins (2007).

In the light of the above, two sets of population were identified to be relevant for this research. One, lecturers in the accounting department and allied departments in the business school of some higher institutions of learning were chosen. They were chosen, being the target users of the ILABS that is meant to generate tutoring systems to support traditional classroom teaching. The involvement of lecturers from other allied departments, outside main accounting department, was undertaken: to benefit from their different perspectives with respect to the ILABS to be evaluated; to access large samples that could be analysed; in realisation that they would have undertaken an accounting course at one time or the other during their studies; and that most of these disciplines also have a numerical aspect that the builder can be extended to, in the near future. Two, undergraduate students taking accounting and finance modules; chosen for being the target users of the tutoring systems that would be authored by lecturers. Preference was given to year-one undergraduate students undertaking the introductory modules because [i] modules that would be evaluated are general to students at this level in most business schools, [ii] this is the entry level for accounting and finance modules, and [iii] the students' population at this level is large, thereby providing a feasible statistically significant sample size that can be analysed.

However, since it was impossible to survey or interview the entire population, sampling remained a viable option to use. Determining the sampling strategy to adopt, due consideration was given to the nature of the research and the target populations. Ultimately, two non-probabilistic sampling strategies, criterion and convenience strategies, were adopted. Samples from both the lecturers and students populations were drawn based on: firstly, criterion strategy—*"choosing settings, groups, and/or individuals because they represent one or more criteria"* (Onwuegbuzie & Collins, 2007, p. 286); and secondly, convenience strategy—*"choosing settings, groups, and/or individuals that are conveniently available and willing to participate in the study"* (ibid).

In essence, the research samples for the questionnaire and interview sessions were based purely on: [i] samples that fall within the research's interest groups or populations—as defined above; and [ii] volunteer participants within the defined population—since they cannot be compelled, and to ensure compliance with ethical rules guiding research involving human subjects. Each population was sampled accordingly, by extending widely a voluntary invitation to members of the targeted population sets. This was undertaken to draw a true representation of the population that would provide information/data leading to valid and reliable generalisation (with respect to the quantitative aspect of the research), and rich data and understanding (with respect to the qualitative aspect). For the survey, the use of the questionnaire technique required a statistically significant sample, as mentioned above. As a result, the guidelines stated earlier were followed (see details in Creswell, 1998; Onwuegbuzie & Collins, 2007). In accordance, a minimum sample size of 102 was targeted, where feasible. While for the interview aspect, a minimum of eight participants (the midpoint of the recommended range—six to ten) was considered. The highlighted figures were only used as a benchmark, to drive the campaign for voluntary participation. A higher number of participants would definitely be considered, because it would enrich the data set, even though it would incur more research time.

Therefore, the combination of questionnaire and interview methods, after taking care of the sampling requirements, provides ground for the emergence of analysable data sets from both the quantitative and qualitative ends. These were analysed and findings triangulated (see triangulation design in Creswell & Plano Clark, 2007, pp.119-120) and interpreted to reach justifiable conclusions. Details on the instruments utilised in this regard are discussed next.

4.4 The Evaluation Instruments

The research methodology comprises two instruments, namely questionnaire and interview protocol (with semi-structured questions). Two sets of questionnaire were developed: the first, addressed question one—propositions 1.1 to 1.4; while the second, addressed question two—proposition 2.1 to 2.4. The first questionnaire was meant to gather data from lecturer users of the ILABS. The second was administered to student users of the modules generated from the ILABS. On the other hand, the interview protocol was only applicable to research question one. The application of the instruments in the research is illustrated below (see figure 4.2).

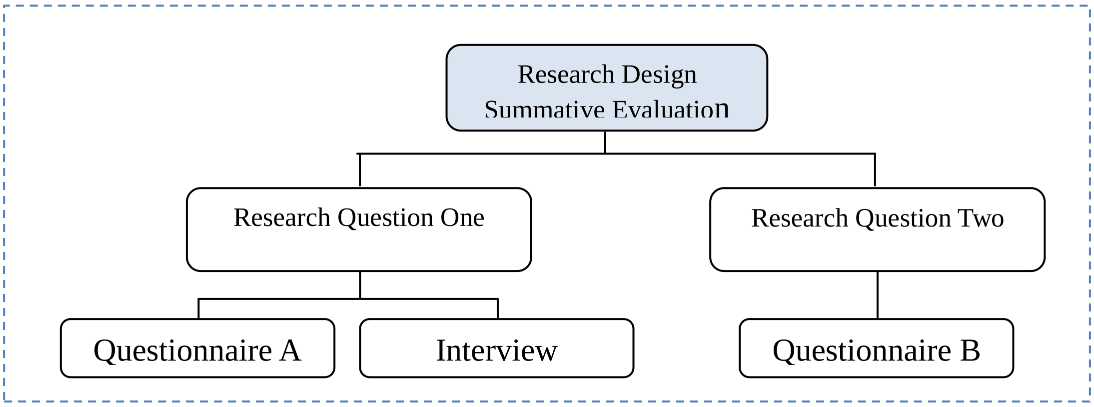


Figure 4.2: Application of Research Instruments within the Research Design

4.4.1 Questionnaires

The two questionnaires utilised in this research were purposely developed, since there was no standard questionnaire(s) that covers all the issues investigated. This covers all the propositions for both research questions, except proposition 1.4 of research question one that addresses usability issues. In that regard, an existing and validated usability questionnaire (QUIS—Questionnaire for User Interaction Satisfaction, developed by Chin, Diehl & Norman, 1988) was adopted. This was undertaken to avoid or limit any subjectivity that the instruments may be subjected to, similar to the approach adopted/adapted in some previous studies (see Moundridou & Virvou, 2001a; Moundridou & Virvou, 2001b, 2003b; Granic, Glavinic & Stankov, 2004; Akilli, 2005; Sykes, 2005; Granic, 2008). So, to develop the questionnaires, each proposition was broken down into its constituent concepts and questions were coined accordingly to arrive at the final instrument (see Gillham, 2000; Frankfurt-Nachmias & Nachmias, 1996).

At the end, each questionnaire had four sections: sections A and B aim to collect participants' details with respect to demographic characteristics and computer experience; section C addresses the respective questionnaire's propositions and contains structured questions using a five-option Likert scale, thus requiring respondents' to select from available options; and section D contains general open-ended or subjective questions—enabling participants to express their views on issues examined, in an economical way and within the tight space(s) provided, in contrast to the unrestricted situation in an interview scene. The first questionnaire, designed to address research question one (propositions 1.1 to 1.4), has 49 items spread across four sections and code named, ***"Builder Questionnaire"*** – (BQ). Similarly, the second questionnaire, meant to investigate research question two (propositions 2.1 to 2.4), has 62 items and is code named, ***"eTutor Questionnaire"*** – (eTQ**)**. Each of the questionnaires has seven scales/constructs respectively (see Table 4.2 and 4.3 below for details).

Table 4. 2: Structure of Piloted Builder Questionnaire

Builder Questionnaire	**Constructs**		**Unscaled Items**	**Scaled Items**	**Open-ended Items**	**Total**
Demographic Characteristics		----	4	---	---	4
Computer Experience		----	4	---	1	5
Proposition 1.1	Builder Assumptions scale	BDASM	---	3	---	5
	Generated Tutors Behaviour	TUTBHV	---	3	---	3
Proposition 1.2	Tutoring Strategies	TUTSTRG	---	3	---	3
Proposition 1.3	Builder Restrictions	BDRST	---	3	---	3
	Production Time	PDTIM	---	2	---	2
	Special Skill	SPSKL	---	2	---	2
Proposition 1.4	Usability	USAB	---	19	---	19
Subjective Items		----	---	---	5	5
Total			***8***	***35***	***6***	***49***

Table 4. 3: Structure of Piloted eTutor Questionnaire

eTutor Questionnaire	**Constructs**		**Unscaled Items**	**Scaled Items**	**Open-ended Items**	**Total**
Demographic Characteristics	----	----	5	---	---	5
Computer Experience	----	----	7	---	1	8
Proposition 2.1	Cognitive Process Visibility	CPVSB	---	7	---	7
	Conversation Aid Cognitive Visibility	CCVSB	---	2	---	2
Proposition 2.2	Timely Feedback	TIMFDBK	---	2	---	2
	Relevance Feedback	RELFDBK	---	7	---	7
Proposition 2.3	Misconception	MISCP	---	4	---	4
Proposition 2.4	Learning Effectiveness	LNEFTV	---	13	---	13
	Cognitive Visibility & Learning	CVSBLN	---	9	---	9
Subjective Items	----	----	---	---	5	5
Total			***12***	***44***	***6***	***62***

However, while developing the questionnaire items, validity (in terms of questionnaire content, construct and criteria) and reliability issues were given particular attention due to their importance (Denscombe, 2007, 2010). In response to that, each question only treats one concept by avoiding a two-in-one question. This was undertaken to ensure reliability of respective scales. Accordingly, Fowler (2002, 2009) stated that, "....

another way to make questions unreliable is to ask two questions at the same time." (p.84; and p.94 respectively) Also, care was taken to draft questions, ensuring preciseness and conciseness, as much as possible. Attempts were made to avoid leading, threatening/emotional wording, double-negatives and double-barrelled questions (Omoteso, 2006; Pallant, 2007), and words with double meanings (Pallant, 2007).

4.4.2 Interview instrument

As per interview instruments, a set of questions was developed to attract in-depth information on issues being addressed in this research. In designing the questions, the research took into cognisance the significant role of questions, i.e. their determinant effect on the outcome of any survey interview. According to Fowler (2009), the most important step in good interviewing is to design a good survey instrument. In order to get quality and a dependable result, there must be good questions. Research on survey instruments shows that certain questions are misread consistently, while others are consistently answered inadequately, thus requiring interviewers to probe in order to obtain adequate answers (Fowler, 1991; Oksenberg et al., 1991; Fowler & Cannell, 1996—all cited in Fowler, 2002). Also, further probing, explanation and clarification of question(s) has been reported as a hindrance to good interviewing. This may present a situation where answers may likely be influenced (Fowler, 2009). Hence, care must be taken to develop questions that eliminate or reduce further probing of interviewees, due to lack of clarity or understanding of the question(s) posed.

In order to achieve the above, the same approach used in developing the questionnaire questions, by addressing one concept or issue at a time, was adopted. All forms of complexity and technicalities were removed, and simple language was used. Both the questionnaire and the interview instruments were validated and their reliability in measuring concepts addressed in the research was tested through a pilot study (discussed in section 4.5 below). The validation and reliability process is detailed in section 4.8 below.

4.5 Pilot Study

A pilot study has been defined in numerous ways. It is understood as a small experiment, designed to test logistics and gather information prior to a larger study, aimed at improving the latter's quality and efficiency (Altman et al., 2006). Equally, the Concise Oxford Thesaurus (Waite, 2002) defined a pilot study as an experiment, test, preliminary trial, or try-out investigation. According to Thabane et al. (2010, p.1), similar definitions are provided in statistics and epidemiology dictionaries, in which they define the pilot study as a small scale...

- *"...investigation designed to test the feasibility of methods and procedures for later use on a large scale or to search for possible effects and associations that may be worth following up in a subsequent largest study" (Everitt, 2006);*
- *"..test of the methods and procedures to be used on a larger scale if the pilot study demonstrates that the methods and procedures can work" (Last, 2001).*

For Emory and Cooper (1991) cited in Omoteso (2006), a pilot study includes (i) pre-testing of research instruments, to detect possible deficiencies in their design and administration, and (ii) clarification of instruments' grey areas that may require further information in order to complete answers to the questions posed. Other definitions are captured in Thabane et al. (2010), and a closer look at them reflects a similar meaning as the ones stated above. In all, they point to a single aim, which is to guide planning and implementation of a large-scale investigation.

4.5.1 Pilot Study of the Questionnaires

Considering the above purposes of a pilot study, the questionnaires were piloted prior to the commencement of the actual research. They were administered on not too divergent targeted subjects that had the characteristics of the focused populations (Omoteso, 2006), but not involved in the main study. This enabled the verification of their validity/reliability. It also helped in reframing/restructuring some questions, ensuring the collection of useful, qualitative and dependable data that was helpful in examining the research issues, thereby leading to meaningful inference and conclusion.

On the first page of each questionnaire, a brief introduction, stating the purpose of the research, and general instruction on how to complete the questionnaire were presented. This included information that participation would be regarded as consent to partake in the research, and participants were assured of anonymity and confidentiality of the information provided and other related issues that may affect their persons. None of the respondents was compelled to participate; it was a wilful decision on the part of these individuals to take part. Also, each questionnaire contains only items relevant to their respective target audience group (i.e. either students or lecturers) (see appendix 4.2 & 4.3). However, those who participated in the pilot study were excluded from the main study in order not to bias findings. Four institutions were involved, and code named to enforce confidentiality (e.g. Uni. A - pilot I only, and Uni. B, C and D—pilot II & main study).

As part of the piloting phase, the questionnaires were presented to experienced researchers to evaluate them. These persons were not included in the actual pilot and main studies conducted. They formed a pre-pilot evaluator, and their comments were as follows:

i. that students may not attend to the subjective section of the "eTutor Questionnaire"—"eTQ"; as such; it should be removed entirely or reduced drastically;
ii. that subjective items should be treated in interview sessions, if feasible;
iii. that students may not understand the technical terms "cognitive process" and "misconception", which were used in the "eTutor Questionnaire" (eTQ);
iv. they observed that there was no subjective section in "Builder Questionnaire"- (BQ);
v. duplicate questions or items were observed in both questionnaires.

In response to the above remarks, the subjective items in eTQ that initially contained 14 items were reduced to five items. On the other hand, the BQ did not contain any subjective items. In order to give respondents opportunity to express themselves, five subjective items were introduced. Also, all technical terms were removed from both questionnaires. For example, the term "cognitive process" was changed to "thinking process"; and "misconception" was changed to "misunderstanding". These responses gave birth to the instruments piloted, structured as earlier shown in Tables 4.2 and 4.3 above.

Accordingly, two pilot studies were conducted. A total of 24 lecturers participated with respect to BQ pilots (pilot I and II), while a total of 43 students participated with respect to eTQ pilots. From the pilot study I, it was clear that the questions were precise and concise; respondents were able to understand them, except three cases of misinterpretation (two cases on eTQ and one case on BQ, as a result of some terms used in the questionnaires. The misinterpretations identified are:

i. the term "Tutor" was sometimes confused to mean "human tutor" in both questionnaires, although it was defined on the first page of each questionnaire;

ii. the terms "Authoring Tool" and "tool" were sometimes taken to mean Tutor;

Based on the stated misinterpretations, both questionnaires were reviewed as stated below:

i. the term "Tutor" was changed to "eTutor";
ii. the terms "Authoring Tool" and "tool" were replaced by the word "Builder";
iii. by eliminating the term "Authoring Tool", the confusion resulting from the word "Tool" was addressed; and
iv. the introduction sections of the two questionnaires were re-phrased to reflect the above changes.

Since the changes did not affect the structure of the questionnaires, only changes relating to the content were affected as identified above (see instruments in appendix 4.2 and 4.3). Hence, the structures of the piloted questionnaires were maintained for the second pilot study. From the second pilot study, no issue was raised that could necessitate changes to the questionnaires. As a result, the questionnaires were administered in the main study.

4.5.2 Pilot Study of the Interview Protocols

Piloting of the interview questions was also undertaken, similar to the questionnaire instruments. A semi-structured interview protocol, aimed for an interview session that should last between 30 and 45 minutes, was developed. This was for lecturers; it was code named "*Builder Interview Procotol"* (BIP). The questions were phrased in tune with the approach posited by David Silvermann (see Silvermann, 2010, p.197). The latter suggested different styles of questioning prior to carrying out the qualitative study. Taking a cue from that, each question was framed in two ways to elicit and compare responses. This resulted in 28 items for the prototype ILABS, being the initial draft (structured as shown appendix 4.4b). This protocol was run through a two-phase critique, involving four (4) experienced researchers, totally excluded from the pilot and main studies. Their responses/observations are noted, thus:

i. the number of questions are too much for an interview session of 30-45 minutes, hence it should be cut down to 10-15 questions, which should include a maximum of three preliminary questions;
ii. some questions are structured, not open enough, and may not elicit an appropriate response, and hence should be reframed; and

iii. technical terms should be avoided, as much as possible, to aid comprehension.

In response to the above, the protocol was reviewed and restructured as shown in Table 4.4 below. The new structure with 15 questions was administered at the pilot interview sessions for the lecturers, to test its effectiveness and the logistics that will be deployed in the final study.

Table 4. 4: Structure of the *Builder Interview Protocol (used for pilot/main study)*

S/N	Headings	Number of Items
1	Preliminary	2
2	Proposition 1.1	2
3	Proposition 1.2	2
4	Proposition 1.3	3
5	Proposition 1.4	2
6	Concluding Questions	4
Total		***15***

Interview questions were sent out, accompanied by a letter of introduction explaining the purpose of the interview. They were notified on the need to make sure that general comments regarding the framing and number of questions sent to them, were appropriate. Each interview session lasted between 30 and 45 minutes, including 10 to 15 minutes demonstration of the target prototype ILABS/tutoring system, except one session which lasted almost two hours. A total of 5 lecturers participated. The involved persons were excluded from the main data collection to avoid biasing the information/data that would be gathered, as argued above. The only observations from the piloted interview sessions relates to two terms, "Tutor" and "Authoring tool". The misinterpretation resulting, were the same as that in the questionnaire. In order to avoid using probe question(s) to clarify main interview questions, they were therefore changed in the interview protocol, thus:

i. the word "Tutor" was changed to "eTutor"; and

ii. "Authoring tool" was changed to "Builder", enabling consistency in the instruments.

After the above observations were effected, the actual structure used in the pilot studies was maintained for the main study undertaken (see table 4.4 above and appendix 4.4a for instrument).

4.6. Data Collection

Data collection was planned in three phases: the first, a pilot study, was undertaken in one university; the second—also a pilot study—took place in three universities; and the third phase was the main study and was equally held in three institutions. During the planning stage of the research, the initial challenge was how to access samples to use in the research. This was resolved through support from colleagues in host institutions (they provided access to their students and fellow lecturers). The data collection phases were not as smooth as would be expected. There were many disappointments, but through persistency and rescheduled visits, it was finally realised. Table 4.5 below shows the phases, activities and the analysis intended per phase. A description of participants involved and procedures taken is also presented.

Table 4. 5: Study Evaluation Phases, Activities and Analysis

Phase	Activity	Analysis
Phase 1: Pilot study I Testing of Questionnaires and Interview protocol	• Exposure of students to generated tutoring systems (eTutor); • Exposure of lecturers to ILABS; • Completion of questionnaire by both students & lecturers; • Interview sessions conducted for lecturers.	• Validation of the exposure logistics with respect to the tutoring systems and ILABS; • Validation of questionnaires used; • Validation of interview protocol.
Phase II: Pilot study II Testing of Questionnaire & Interview protocol	(same as above)	(same as above)
Phase III: Main Study Main Evaluation using Questionnaires & Interview Protocol	• Exposure of students to tutoring systems at three universities • Exposure of lecturers to ILABS at three universities; • Completion of questionnaires by both students & lecturers; • Conduct of interviews for lecturers.	This phase provides insight into the research questions posed. Findings with respect to the questions can be found in chapters 5 and 6 of this thesis.

4.6.1 Participants

This research involved both student and lecturer users of the generated tutoring systems and the prototype ILABS respectively. With respect to students, a total of 300 completed questionnaires were retrieved from students, who participated in the main study. On the other hand, several lecturers in accounting and finance department and other allied departments in the business schools of host institutions were contacted. As a result, a total of 82 completed questionnaires were retrieved from lecturers and eight interviews were conducted; both with respect to the main study.

4.6.2 Data Collection Procedure

In order to access data from both students and lecturers, an ethical approval form (see appendix 4.1) was completed. The consent of heads of various departments in host institutions was secured through colleagues, who provided the link. Thereafter, arrangements were made with lecturers to solicit their participation and to provide a link to their students as well. Lecturers who showed interest, were contacted for exposure sessions to provide hands-on experience on the ILABS and the tutoring systems generated, questionnaire administration and to arrange interview appointment dates/times. Letters were sent out a week before commencement of the evaluation. These introduced the purpose of the evaluation with copies of the research instruments. After exposure sessions, subjects completed the questionnaire containing mainly closed-questions and very few open-ended questions.

Thereafter, interview sessions were organised on a private basis. This was held in the individual lecturer's office space, and conducted at their convenience, considering their busy schedules. The schedule was followed strictly, with a date/time slot taken by each respondent, as booked, and took two weeks to complete. Interview questions were semi-structured, tailored along the propositions, which were being examined. These sessions were digitally recorded with permission of the interviewees, to ensure no loss of information and ascertain their willingness to participate. Also, the number interviewed was pre-determined, taking into consideration factors, such as (1) time, and (2) suggested minimum interviews required for a qualitative study, which ranged between six and ten, as earlier discussed above (see Collins, Onwuegbuzie, & Jiao, 2007; Onwuegbuzie & Collins, 2007). Accordingly, not all participants that took part in the questionnaire, volunteered to be interviewed.

With respect to students, the lecturers contacted provided access to their students. These students were addressed after their lectures in host institutions to solicit their participation. None of the students were compelled in any manner. Participation was made optional and was distinctively clarified in an open speech made in the classroom. Similar to the procedure adopted for lecturers, exposure sessions were organised to provide hands-on experience required to give a feel for the tutoring systems and identify their characteristics. After period of exploration, subjects were made to complete a questionnaire containing mainly closed-questions and very few open-ended questions.

4.7 Data Analysis

As a result of the data collection, two types of analysis were employed, quantitative and qualitative analyses, to address quantitative and qualitative related data respectively.

4.7.1 Quantitative Analytical Procedures

As a first step towards analysis, data from the two questionnaires were entered into the IBM SPSS statistics version 19, followed by screening. This involved checking data entry errors, missing values and outliers; where errors existed, they were removed. This was implemented by generating frequency statistics of all variables, which showed the frequency of valid cases, missing values, and minimum and maximum values—in order to determine whether the values fell within the expected range. Missing values were cross-checked with original documents to ascertain whether they were genuinely missing. The report, therefore, shows no error; neither were there out-of-range values, and missing values were confirmed real. Missing values were very few, in this instance below the suggested 5% upper boundary (Tabachnick & Fidell, 2007) in case(s)/variables where they occur.

In order to address missing values, the pairwise deletion technique was employed during analysis, because it excludes case(s) only if there is missing data required for a specific analysis, unlike "listwise" deletion approach that includes only cases with data on all the variables constituting each case, irrespective of the analysis embarked on (Pallant, 2010). Consequently, the latter approach tends to lose quite a number of cases, thereby drastically reducing the total sample size. This may impact negatively on some analyses. The former approach limits the number of cases removed. It helped eliminate or, at least, reduce any effect, if any, that may impact the sample size.

Subsequently, various analyses were employed to explore the data sets. Categorical variables, which include the demographic characteristics and computer experience variables, were analysed using frequency analysis. This was undertaken to determine the distribution of participants across segments of each variable. Thus, variables with evenly distributed participants were determined to enhance their use in subsequent investigation of continuous variables (i.e. questionnaire scales/constructs). On the other hand, continuous variables, i.e. scales/constructs and their items, were analysed using descriptive statistics, i.e. mean, median, standard deviation, skewness and kurtosis, and Bi-variate statistics, i.e. t-test, Analysis Of Variance (ANOVA) and Correlation.

The application of descriptive statistics enabled the determination of participants' views with respect to the constructs measured. This was achieved by comparing the mean scores of participants with the research benchmark (3.0)—the mid-point of the five-point Likert scale used in the research. This benchmark was purposely chosen due to the nature of the evaluation (in which there is no previous version to use as benchmark), and in view of the fact that a similar approach was used in the literature (see Granic, 2008). In the latter work, a benchmark was chosen via a pilot study. However, in this research, the benchmark was chosen based on the five-point scale used. Specifically, the middle point (i.e. 3.0), which represents "neither agree nor disagree point" was utilised as the benchmark in the analyses. Consequently, mean scores below the benchmark indicated disagreement to the construct(s) measured, while the mean score above the benchmark indicated agreement.

In addition, Bi-variate statistics enabled further investigation of participants' views in order to determine factors that might have contributed to users' reactions or opinions on constructs. This was realised by comparing mean scores of constructs across categorical variables segments. The only exception was correlation statistics, which was specifically employed to examine the existence of a relationship between two constructs examined in the research. Table 4.6 below shows a summary of the statistical techniques that were utilised in the research. Also, appendix 5.1 provides a detailed discussion of the preliminary analysis undertaken, using the stated statistical techniques, with respect to data collected in this research (i.e. the three phases).

Table 4. 6: Statistical analysis for students and lecturers data sets

Data Type	Proposed Analysis	Reasons
Questionnaire-closed items (students & lecturers data)	• Descriptive statistics	• In order to ascertain accuracy of data entered. • Determine the normality of questionnaire scales (a requirement that must be ascertained before parametric techniques can be used). • Determine overall users' views/reactions regarding the constructs measured.
	• Reliability test	• Determine if the scales items measures supposed constructs.
	• One-sample T Test	• Compare the users' perceptions against the research benchmark in order to determine agreement or disagreement with the construct measured.
	• Independent-samples T test • One-way ANOVA	• Compare the users' perceptions across various categories (e.g. universities, departments, gender etc.)
	• Spearman correlation	• In order to examine the relationship between two constructs (BDASM & TUTBHV).
Questionnaire open-ended items (students & lecturers data)	• Extract and categorise themes	• Provides a qualitative perspective of respondents' views.

4.7.2 Qualitative Analytical Procedures

With respect to the qualitative analysis, data from the interview sessions were transcribed, coded, and analysed using inductive and deductive thematic analysis in consideration of the literature (Miles & Huberman, 1994; Braun & Clarke, 2006; Fereday & Muir-Cochrane, 2006; Auld et al., 2007; Burnard et al., 2008; McMillan, 2009; Costu, Aydin & Filiz, 2009; Cruzes & Dyba, 2011; Chenail, 2012). Thus, the research employed two layers of analysis: the inductive stage—which enabled themes to emerge in a naturalistic way; and deductive stage, which was used to screen/categorise themes according to the pedagogic metamodel that underpins the research. This draws on previous studies that demonstrated a similar procedure (Fereday & Muir-Cochrane, 2006; Burnard et al., 2008; Cruzes & Dyba, 2011).

The analytical process involved transcription of each interview, with subsequent digitalisation using Microsoft word. Thereafter, transcripts were checked and re-checked several times, read word by word while listening to the digital recorder, to ensure no errors. This was followed by several coding and re-coding iterations to generate first level codes. Coded transcripts were then merged into one Microsoft Excel file for further coding levels and categorisation of similar codes. The coding and re-coding process equally involved four other persons, who helped validate (i.e. peer review) the codes generated. This was undertaken to ensure rigour, and credibility and conformability of the analysis, and to eliminate or reduce bias (Fer, 2004; Auld et al., 2007; Burnard et al., 2008; Cruzes & Dyba, 2011). Accordingly, all the four persons agreed to the codes generated, except two duplications that were identified and addressed immediately.

Flowing from the above analytical approach, an understanding of the meaning of the themes that emerged from the codes was achieved. The themes drawn from codes were then screened along the theoretical framework used in the research. Consequently, key findings under each theme were presented using descriptive and interpretive reporting methods, which have been used or mentioned in previous studies (Fer, 2004; Burnard et al., 2008). Thereafter, findings from the qualitative analysis were integrated with that of quantitative analysis, where applicable (since some research issues were only investigated quantitatively). The integrated findings were then discussed in the light of the previous studies in the literature (Burnard et al., 2008). This enabled the confirmation or refutation of claims made with respect to the ILABS, and its underlying theoretical assumptions, and other issues examined in the research. Also, deeper insights were derived from the qualitative aspect, and it enhanced the identification of some issues not envisaged in the research.

4.8 Validity/Reliability Considerations

Validity and reliability are regarded as key elements/indicators of research quality. As such, the research instruments/procedures employed should be ascertained for data collected to be accepted as a true measure, replicable, and upon which meaningful interpretation could be made (Kimberling & Winsterstein, 2008; Fowler, 2009). Defining these indicators, Bennett (2003) noted, in terms of validity:

> *...data are said to be valid if they measure what they claim to be measuring" (p. 100); a*nd with respect to reliability, *"....data are said to be reliable if repeating the technique gives the same result again. (p. 98)*

Based on the above, Pallant (2007) suggested that the two types of reliability, test-retest (also known as temporary stability) and internal consistency, should be carried out. The first determines if the instrument yields the same result on replication. However, this also depends on the nature of the scale being measured. For scales that are time/situational bound, e.g. measuring mood, this would not be appropriate. The second examines whether the scale items hang-up, i.e. they measure the same construct. Due to their significance, therefore, it is important to undertake trials of a research instrument in order to ensure validity and reliability of its measures (Kimberling & Winsterstein, 2008). Also, Holliday (2007) claimed that different sources of validity associated with quantitative and qualitative research exist. On the former, the research should report detailed procedures undertaken, while for the latter, a demonstration of the appropriateness of the overall strategy employed in the social setting, the researcher-subject relationships within it, and the steps taken for thorough engagement, should be explained.

While the above discussed concepts―validity and reliability―are predominantly applied in a quantitative paradigm, there were varied positions on their application in qualitative research (Ali & Yusof, 2011). In recognition of the arguments for and against the notions of validity and reliability in qualitative research, Janesick (1994) argued that these concepts should be applied to all research—quantitative and qualitative (cited in Ali & Yusof, 2011). In that light, some researchers fail to mention them in their research, while some used different terms instead, such as credibility, rigour, conformability, trustworthiness etc. (Fer, 2004; Ali & Yusof, 2011). As a result, this thesis adopted some terms defined by Lincoln & Guba (1989): “credibility” and “conformability” in place of validity and “dependability” rather than reliability—as cited and utilised in Fer (2004), with respect to the qualitative aspect of this research.

In the light of the above, the questionnaire items were developed according to each research proposition, except with respect to the proposition on usability of the ILABS, in which the research utilised a standard and validated instrument. Despite that, the instruments were peer-reviewed by four experts, and changes effected as identified (see section 4.5.1). Also, the instruments were piloted twice on students and lecturers, who are not very divergent from the real subjects considered for the main study. Details of the process, including data and analytical procedures, sampling scheme and size, and statistics employed were discussed and justified in earlier sections above. Furthermore, reliability analysis of the questionnaire scales was undertaken using IBM SPSS statistics version 19 to ascertain the reliability of the scales. This was measured using Cronbach alpha coefficient benchmark (0.7) as recommended in the literature (De Vellis, 2003 cited in Pallant, 2007). For scales yielding low alpha value, and with small number of items (i.e. less than ten), the mean inter-correlation was computed to determine the strength of the relationship among the items of the scale (Pallant, 2010). In relation to this research, each of the scales had less than ten items, except the usability scale, but the result shows a strong relationship among items in each scale. See appendix 5.1 (section 5.32) and 6.1 (section 6.3) respectively for detail results of the reliability test with respect to the two questionnaires' scales.

On the other hand, the interview protocol was developed in response to each of the propositions, peer-reviewed and piloted as undertaken in the quantitative aspect of the study. In order to ensure credibility and conformability of the qualitative aspect, the rationale and process was explained (see sections 4.3 to 4.5 above). Also, sampling scheme and size were chosen in conformity with previous studies. Transcribed interviews were checked several times, line by line and word by word, while listening to the digital recorder. Several iterations of coding, re-coding, and categorisation of similar codes was undertaken to ensure true representation of the data. Also, four researchers, external to this research, were employed to check codes generated from the qualitative data. On dependability, the detailed analytical procedures undertaken were provided in the above identified sections to enhance judgment and limit any bias that research may be subjected to.

4.9 Limitations of the Methodology

As described above, this research utilised both quantitative and qualitative research approaches to examine the research questions, where applicable. Thus, it draws on the inherent advantages that accrue from the application of both approaches, as mentioned in various pieces of literature (Gillham, 2000; Bennett, 2003; Tashakkori & Teddlie, 2003), eliminating, or at least reducing, the individual weaknesses of either approaches (Gillham, 2000; Bennett, 2003; Silverman, 2010). According to Gillham (2000), *"...different methods have different, even if overlapping, strengths and weaknesses. If you use a range of methods you can put together a more adequate picture." (p.81)* So, this research attempts to take advantage of the aforementioned in order to arrive at reliable findings that could provide deeper insights, as well as confirm or refute claims made in this research.

Despite the above, some limitations could be associated with the implementation of the methodology in this research. These could have been eliminated (or reduced) were there enough resources in terms of time, financial strength, and access to a higher number of research subjects. Since this research is being carried out as part of a doctoral programme, resources available are very limited; thus limiting what could actually be achieved within available resources. For instance, interviews could not be undertaken with respect to research question two. The research could have benefited, but was constrained. Notwithstanding, the findings were not in any way invalidated.

Also, experimental research strategy, common in education and psychology—two disciplines contributing to the ITS/Authoring research field, could have been included in the methodology. This research strategy has been identified as suited for educational systems in that it enables investigation of the relationships between teaching interventions and student-related teaching outcomes; it obtains quantitative measures of the significance of the relationships; and often used in summative evaluation, where formal power and conclusions are desired, rather than acquisitions of information (Mark & Greer, 1993). Other research strategies that could be employed include, for example, action research. At an early stage, both research strategies were conceived as possibilities that could enrich data collection and findings, but implementing them was not feasible, given that they are time consuming and expensive, which this research cannot accommodate. They could deepen investigation of research issues, as well as open up hidden issues not considered at the outset of a research. This could contribute to the improvement of the ILABS developed and tutoring systems generated from it.

Adequate resources were not available to carry out an empirically-sound formative evaluation. So, due to time restrictions and other limitations, formative evaluation could not be done in the first instance. Thus, the research settled for a summative evaluation. It is assumed that in a future extension of this work, the outcome of the summative evaluation in this research could be transformed into formative results to improve tools developed/generated. Further, formative evaluations may then spring up from that point, rounded up with a final summative evaluation that attests to (or refutes) claims made with respect to the system and its theoretical assumptions.

4.10 Summary

In this chapter, effort was made to throw light on the evaluation of the methodological approaches adopted, as well as describing the evaluation protocol implemented. Also, previous evaluation studies in the ITS/Authoring field were referenced to corroborate the stance taken in this research. The techniques used to collect data, mainly questionnaire (for the quantitative data), and interview (qualitative data) were discussed and justified. This evaluation study further augments previous studies in the field. It shows further how quantitative techniques combined with qualitative technique can be used to gather a useful and meaningful data set, in an attempt to explain educational issues emerging from technology-enhanced tutoring systems.

The next two chapters describe the empirical aspect of this research. Data collected was analysed, using the statistical techniques identified in the current chapter. The analysis centred on the prototype ILABS and tutoring systems generated from it. Findings were discussed in relation to issues investigated.

Chapter 5: Data Analysis I

This chapter provides answers to research question one. It presents both the quantitative and qualitative analyses of the data collected with respect to the prototype ILABS. Mixed methodology was employed due to benefits advanced earlier (see chapter 4); these include complementary advantage of the approaches involved (Pascal—cited in: Tashakkori & Teddlie, 2003; Fer, 2004) and concurrent validity that ensues (McMillan 2009)

In consideration of the aforementioned, an attempt was made to gain new and deeper understanding from data provided by participants through quantitative (or probabilistic) and descriptive interpretations, aimed towards addressing some research objectives stated earlier in chapter one. The process taken to realise the above is stated accordingly below.

5.1 Data Collection/Analysis

In order to realise the empirical aspect of research question one, data collection stretched through three phases, as stated in the previous chapter. Two phases were pilot studies, while the third phase was the main (or actual) study. The conduct of two pilot studies enabled the testing of the research instrument, to evaluate its reliability and validity. However, in this chapter, the analysis of the main study data (i.e. third phase) is presented, while the analysis of the pilot studies can be found in appendix 5.1. Accordingly, two forms of analysis are presented, quantitative and qualitative. Each employing a relevant analytical procedure, as previously discussed, in an attempt to address the four research propositions of research question one.

5.2 Research Instruments / Participants

In accordance with above, the instruments listed below (see figure 5.1) were administered to university lecturers in business and allied courses, being the targeted population (the contemplated users of the ILABS developed). The instruments were administered in three higher education institutions (code named: Uni. B, C and D) that participated in the main study.

Instruments:

- Builder Questionnaire (lecturer users only)
- Builder Interview protocol (lecturer users only)

Figure 5.1: The Research Instruments

Participants were drawn from key relevant departments in the business schools of the institutions concerned, mainly accounting, and banking and finance, although other departments were included. The reason was that the context of the research focused on numerical problem solving, which includes accounting and finance modules. So, lecturers within these disciplines were considered appropriate for this research. Table 5.1 below provides the statistics of participants. Note that code-naming of institutions, as reflected in the table below, was done for confidentiality purposes.

Table 5. 1: Number of Participants in the Main Study

Study Type / Institutions		**Questionnaire**	**Interviews**
Main study	Uni.B	32	5
	Uni.C	25	---
	Uni.D	25	3
Total		**82**	**8**

5.3 Quantitative Analysis

This section presents the quantitative analysis of the research, mainly analysis of responses to the questionnaire instrument. Mertlier & Charles (2005) and Pallant (2010) provided clues on the types of quantitative analyses that can be employed. According to them, prior to extensive analysis that may involve lots of energy, time, and other research resources, data should be treated to certain preliminary analysis. It involves checking and cleaning data entered into a statistical package, purposely, to eliminate (or at least, reduce drastically) data entry errors. This is a crucial analytical step, because many statistics are sensitive to slight change due to data capture errors (Pallant, 2010), so allowing it may result in a false outcome. It will also identify, where present, missing data. Equally, many of the statistics have certain assumptions that should be met before they can be applied. For example, parametric statistics (e.g. t-test, ANOVA, correlation etc.) require normally distributed data. This then requires confirmation of the nature of the data before deciding whether to apply such a class of statistics or not.

In the light of the above, certain steps were taken prior to the main analysis. This includes data cleaning, factor analysis—where applicable, reliability test and descriptive statistics. The statistics enabled the testing of statistical assumptions, verification of the data entered into IBM SPSS Statistics version 19 and exploration of their nature, prior to further analyses. Through these means, all forms of data capturing errors were identified and eliminated to prevent negative influences on the result used in the core analysis of the research. Moreover, the steps informed the statistics that were eventually used in the final analysis. Details of the preliminary analytical procedures can be found in appendix 5.1, while the main quantitative analysis is presented below.

As stated in chapter one, research question one was broken down into four propositions to enable investigation. Each proposition was treated to one or more quantitative analyses—as may be required, drawing insights from users' views/reactions to research objects being evaluated (i.e. prototype ILABS and ILAT). Also, it aimed to determine the respondents' perceptions of the theoretical constructs being examined. Findings were discussed in the light of theoretical framework underlying this work, the research context, previous studies and practice in the ITS/Authoring field.

In line with the foregoing, propositions one and two intend to capture users' opinion on the extent to which the prototype ILABS aligns with its underlying theoretical constructs. Proposition three checks for the presence of some key features in the builder, while proposition four addresses the usability of the system in general. Since there are no previous versions of the system being evaluated, the research is limited in terms of the nature of analysis that can be undertaken. So, achieved mean scores were compared against a pre-set benchmark using both descriptive statistics and one-sample t-test. The use of a pre-set benchmark draws from previous work in the literature (see Granic, 2008). In the latter work, the benchmark was chosen via a pilot study. However, in this research, the benchmark was chosen based on the five-point scale used. Specifically, the middle point (i.e. three), which represents "neither agree nor disagree point" was utilised as the benchmark in the analyses undertaken. The comparison is aimed at determining the direction of users' reaction/perception on constructs measured.

Equally, the use of a one-sample t-test aimed to investigate the existence of a statistically significant difference between the predefined benchmark and the compared mean reaction. This was intended to enhance the conclusion made with respect to users' reaction/perception on various scales. Where significance occurs, further investigations were made to determine factors that might have contributed to such difference. These factors only include demographic and computer variables, which are analysed in the preliminary analysis (see appendix 5.1). Specifically, they are institution, department, gender and previous experience with an authoring tool. With respect to the first two, one-way analysis of variance was utilised, because there were three groups in each. On the other hand, the independent t-test was applied to the last two variables, because each had two groups respectively.

However, only the main effect of the stated factors was conducted. The interaction effect of those factors could not be done in this aspect of the research, because there was no sufficient sample size per group to achieve 0.80 powers for 0.05 alpha level—the chosen power/significance level, unless otherwise stated. Power refers to the probability of rejecting a null hypothesis when the alternative hypothesis is true (Tabachnick & Fidell, 2007). In order to detect the above stated power, the number of samples that will be required per group will depend on intended effect size and the number of groups involved (see Cohen, 1992, p.158). When the interaction effect is studied, more groups are involved, and the number of samples that would be required will rise. This was not realisable in the research, since the total sample size involved was moderate (82), while the distribution of the samples across factors did not enable such a level of analysis.

Based on the above, the following sub-sections present analysis of the propositions in a view to answer research question one:

> *What is the perception of users on the conception and implementation of a metamodel in ILABS, which can be utilised by non-programmers to generate an unrestricted number of tutoring systems in a numerical problem-solving context of applied numerical domains?*

5.3.1 Reaction Evaluation of Builder Theoretical Constructs

This section treats the first two propositions that address the core theoretical constructs, thus: [i] *a metamodel can be conceptualised and implemented in ILABS; and* [ii] *It is possible to generate tutoring systems that support process monitoring and model-tracing from the implemented metamodel (i.e. ILABS).* Three out of seven scales that constitute the builder (i.e. ILABS) questionnaire instrument address the two propositions in question, builder assumption scale, generated tutor behaviour, and tutoring strategy scales. Details of the items that constitute the stated scales are provided in table 5.2 below, while table 5.3 provides some basic statistics on users' reaction to the scales. Included in table 5.3 are means reaction to each of the three scales, as well as their individual items. Detailed analysis of the two propositions, based on the table data, follows.

Table 5. 2: Description of Scale Items

Builder Assumption scale (bdasum):

1	bdasum1	The builder system gives me the option to produce eTutor(s) that enable interactive learning.
2	bdasum2	The builder system gives me the option to generate eTutor(s) that adapts feedback to students' thinking processes
3	bdasum3	The builder system allows me to produce eTutor that enables interactive learning as well as adapts feedback to students' thinking processes

Generated Tutor Behaviour scale (TUTBHV):

1	tutbhv1	Response from generated eTutor reflects student's learning process.
2	tutbhv2	The generated eTutor enables interaction between learner and the system.
3	tutbhv3	The generated eTutor monitors learner's problem solving steps.

Tutoring strategy scale (tutstrg):

1	tutstrg1	The builder system allows me to generate eTutor that monitors student's problem solving steps.
2	tutstrg2	The builder system allows me to produce eTutor that traces student's input value.
3	tutstrg3	The builder system can produce eTutor(s) that support both features (i.e. monitors problem solving steps and traces student's input value).

***Note:** The term "generated tutor" or "tutor generated" or "eTutor" refers to ILAT generated from ILABS

Table 5. 3: Statistics of Users Reaction in order to Build Theoretical Constructs

Scale / Scale Items	N	Minimum	Maximum	Mean	Median	Std. Deviation
Builder Assumptions scale	**82**	**4.00**	**5.00**	**4.54472**	**4.66667**	**.410801**
bdasum1	82	4	5	4.65	5.00	.481
bdasum2	82	4	5	4.50	4.50	.503
bdasum3	82	4	5	4.49	4.00	.503
Generated Tutor Behaviour scale	**82**	**4.00**	**5.00**	**4.52439**	**4.66667**	**.395553**
tutbhv1	82	4	5	4.40	4.00	.493
tutbhv2	82	4	5	4.67	5.00	.473
tutbhv3	82	4	5	4.50	4.50	.503
Tutoring Strategy scale	**82**	**4.00**	**5.00**	**4.56098**	**4.66667**	**.456257**
tutstrg1	82	4	5	4.56	5.00	.499
tutstrg2	82	4	5	4.59	5.00	.496
tutstrg3	82	4	5	4.54	5.00	.502

5.3.1.1 Proposition 1.1

A metamodel can be conceptualised and implemented in ILABS

This proposition has both empirical and non-empirical dimensions. The non-empirical dimensions, the conception and practical implementation of a pedagogic metamodel in a prototype ILABS were discussed in chapter 3. However, since the prototype is a practical representation of the metamodel (in terms of its features), an empirical confirmation of the presence of the theoretical assumptions becomes necessary. Here, an attempt is made to address the empirical aspect only, to determine the alignment of the prototype ILABS to three key theoretical assumptions of the metamodel:

i. that learning takes place through conversations;
ii. that the cognitive process can be made visible to aid learning; and
iii. that the cognitive process can be made visible through conversations between learner and domain expert or tutoring system. The latter assumption integrates the first two.

In order to address these assumptions, which the proposition is seeking to do, some questions come to mind, thus:

- Does the ILABS support features that enable the generation of tutoring systems (i.e. ILAT) underpinned by the implemented metamodel assumptions within the numerical problem-solving context of the accounting and finance discipline?
- Is there a difference in the opinions of users across demographic characteristics and computer experience?
- Is there any relationship between the assumed theoretical construct of the ILABS and the generated ILAT?

Answers to these questions are meant to address the empirical aspect of the proposition, determined by seeking to know users' reactions/views on builder questionnaire scales after formal exposure to the prototype ILABS. In this particular regard, two scales are relevant: builder assumptions scale ("BDASUM"—an Independent Variable, "IV"), and generated tutor behaviour scale ("TUTBHV"—a Dependent Variable, "DV").

■Investigating Users Reaction/Perception—Univariate Analysis of Scale/Scale Items

From table 5.3 above, scores on "BDASUM" and "TUTBHV" scales and for their items, range between 4 (agree) to 5 (strongly agree). The overall mean reaction, median and standard deviation for respective scale are: “BDASUM”—4.54, 4.67 and 0.41 respectively; and “TUTBHV”—4.52, 4.67 and 0.396 respectively. The trend is also reflected in the scales’ items; their scores range from 4 to 5, and none of their mean and median scores were below 4.0. On a general note, overall mean reaction of the scales, as well as the means reaction of the scales items were above the benchmark (3.0) utilised. This demonstrates users’ affirmation of the implementation of the theoretical assumptions in ILABS, as well as in the ILATs generated.

However, at a more detailed level, it could be observed that item one (bdasum1—signposting conversation) had the highest mean reaction (4.65) compared to the other scale items, while item three (bdasum3—signposting both conversation and cognitive visibility) had the lowest mean of 4.49. Although the three means suggested strong views in favour of the assumptions, the variations in the means indicated slightly stronger feeling in favour of interaction learning (i.e. conversation assumption—bdasum1) compared to cognitive visibility assumption (represented by bdasp2 scale item). This informs thinking along the lines that ILABS may be more capable of generating ILATs that support conversations between students and the expert system, than support for cognitive process visibility. An inspection of the corresponding item in the generated tutor behaviour scale, revealed the same trend; tutbhv2—signposting conversation assumption—had a mean (4.67) greater than means of the other two scale items, tutbhv1 (4.40) and tutbhv3 (4.50)—both measuring cognitive visibility. Hence, the results seem to strengthen the conclusion that users feel more strongly in favour of conversation than cognitive visibility. The conversation assumption seems slightly more practically felt by users than cognitive visibility. Nevertheless, reaction to both scales suggests affirmation of the implementation of the theoretical assumptions in both the ILABS and the generated ILAT.

■ **Investigating Significant Difference between Benchmark and Scale/Scale Items**

Despite the above, further investigation was undertaken using a one-sample t-test. The means overall reaction to the scales/items were compared with the benchmark, to enable confirmation or refutation of the above stated outcome/conclusion.

The results of the t-test are presented in table 5.4 below and it shows the mean difference for respective items. This represents the difference between the benchmark value and the means of the scales/scales items. In order to interpret therefore, eta squared was computed using the formula suggested in Pallant (2010), thus:

Eta squared = t^2 / (t^2 + (n -1)), where t = t-test value and n = the number of cases.

Figure 5.2 Eta squared formula
Source: Pallant (2010)

Usually, eta squared is meant to explain the proportion of variance in a dependent variable that is explained by an independent (group) variable (ibid). In this specific context, it was used to explain the magnitude of the difference between the benchmark value and mean(s) reaction. In the referred context, the computed eta squared value, which should range between 0 and 1 (ibid; Tabachnick & Fidell, 2007), was interpreted using the guidelines proposed by Cohen (1988, pp. 284-7), thus: 0.01 = small effect; 0.06 = moderate effect; and 0.14 = large effect.

Table 5. 4: One-Sample t-Test for BDASUM and TUTBHV scales

	Test Value = 3.0 (benchmark)					
					95% Confidence Interval of the Difference	
	T	Df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Builder Assumption scale	**34.051**	**81**	**.000**	**1.544715**	**1.45445**	**1.63498**
bdasum1	30.991	81	.000	1.646	1.54	1.75
bdasum2	27.000	81	.000	1.500	1.39	1.61
bdasum3	26.788	81	.000	1.488	1.38	1.60
Generated Tutor Behaviour scale	**34.898**	**81**	**.000**	**1.524390**	**1.43748**	**1.61130**
tutbhv1	25.739	81	.000	1.402	1.29	1.51
tutbhv2	31.996	81	.000	1.671	1.57	1.77
tutbhv3	27.000	81	.000	1.500	1.39	1.61

Based on the above explanation, the results of table 5.4 above indicate that there is significant difference—at 0.05 alpha level—between the benchmark value (i.e. test value) and the means of the scales, thus:

- builder assumption scale (BDASUM)—t (81)=34.051, p=0.0; the magnitude of the difference (mean difference=1.54, 95% CI: 1.45 to 1.63) was very large (eta squared=0.93).
- generated tutor behaviour (TUTBHV)—t (81)=34.898, p=0.0; the magnitude of the difference (mean difference=1.52, 95% CI: 1.44 to 1.61) was very large (eta squared=0.94).

Due to the existence of significant statistical difference and given its achieved magnitude, it can be concluded that users strongly affirm the presence or implementation of the theoretical constructs in ILABS. Hence, the first question raised with respect to this proposition was hereby answered in the affirmative.

■ Investigating the Main Effect of Demographic & Computer Experience Factors

Since a statistically significant difference exists as confirmed above, investigation of the main effect of demographic and computer experience factors on users' reaction—as raised in the second question with respect to this proposition—becomes necessary. This was intended to provide a clue to, or detailed explanation of, the judgement of users. The factors investigated comprise institution, department, gender and previous experience with an authoring tool.

[a] Effects of Institution and Department Factors:

With respect to institution and department factors, Lavene's test of homogeneity of variances across respective groups was not significant for both scales; thus, for institution factor (builder assumption scale—F=0.951, Sig.= .391; and generated tutor behaviour—F=1.134, Sig.= .327); while for department (builder assumption scale—F= 0.266, Sig.=0.767; and generated tutor behaviour scale—F=2.330, Sig.=0.104). As a result, the assumption of homogeneity of variance was not violated, giving credence to the applicability of one-way analysis of variance (ANOVA) in this section of the work. The result of the one-way ANOVA that examined the effect of the three institutional groups on builder assumption and generated tutor behaviour scales (see tables 5.5 below), did not reach statistical significance at 0.05 alpha level: BDASUM—F (2, 81)=0.905, p=0.409; and TUTBHV—F (2, 81)=0.401, p=0.671.

Table 5. 5: one-way ANOVA on Effect of Institution on BDASUM & TUTBHV scales

		Sum of Squares	df	Mean Square	F	Sig.
Builder Assumption scale	Between Groups	.306	2	.153	.905	.409
	Within Groups	13.363	79	.169		
	Total	13.669	81			
Generated Tutor Behaviour scale	Between Groups	.127	2	.064	.401	.671
	Within Groups	12.546	79	.159		
	Total	12.673	81			

Similarly, results of one-way ANOVA of department effect on both scales (see table 5.6), did not reveal statistical significance at 0.05 alpha level: BDASUM—F (2, 81)=0.109, p=0.897; TUTBHV—F (2, 81)=1.391, p=0.255. Therefore, it can be concluded that both institution and department factors did not influence the reaction/perception of users on the respective constructs measured.

Table 5. 6: one-way ANOVA on Effect of Department on BDASUM & TUTBHV scales

		Sum of Squares	df	Mean Square	F	Sig.
Builder Assumption scale	Between Groups	.037	2	.019	.109	.897
	Within Groups	13.632	79	.173		
	Total	13.669	81			
Generated Tutor Behaviour scale	Between Groups	.431	2	.216	1.391	.255
	Within Groups	12.242	79	.155		
	Total	12.673	81			

[b] Effect of Gender and Authoring Tool Experience:

Investigation of the main effect of gender and authoring tool experience on the above scales, revealed that Lavene's test of equality of variances was not significant for both factors with respect to the two scales; i.e. gender (BDASUM: F=0.149, Sig.=0.70; TUTBHV: F=1.442, Sig.=0.233); and authoring tool experience (BDASUM: F=3.382, Sig.=0.07; TUTBHV: F=1.080, Sig.=0.302)—see appendix 5.4a and 5.4b respectively for relevant tables. On effect of gender, results of the t-test at 0.05 alpha level, corresponding to equal variance assumed, did not attain statistical significance for both the builder assumption scale as well as the generated tutor behaviour scale. Indeed, BDASUM: t (80)=1.635, p=0.106, and the magnitude of the difference (mean difference=0.161, 95 CI: -0.035 to 0.358) was small (eta squared=0.03); TUTBHV: t (80)=0.766, p=0.446, and the magnitude of the difference (mean difference=0.074, 95 CI: -0.118 to 0.265) was very small (eta squared=0.007).

Unlike gender factor, the effect of authoring tool experience on builder assumption scale, examined using independent samples t-test, showed a statistically significant difference at 0.05 alpha level, with the "Yes" group (M=4.44, SD=0.418) and "No" group (M=4.68, SD=0.360); t (79)= -2.803, p=0.006. Despite the significance achieved, the magnitude of the difference (mean difference= -0.245, 95 CI: -0.419 to -0.071) was moderate (eta squared=0.09). However, similar to gender factor, the effect of authoring tool experience on generated tutor behaviour scale did not achieve statistical significance, with the "Yes" group (M=4.49, SD=0.377) and the "No" group (M=4.58, SD=0.413); t (79)= -0.958, p=0.341. The magnitude of the difference (mean difference= -0.084, 95% CI: -0.259 to 0.091) was small (eta squared=0.01). Despite the range of effect size, from small to moderate as per authoring tool experience, the pattern of the reaction seems to be similar. The overall reaction means of users who had previous authoring tool experience was consistently lower to those without previous experience for both scales; even though the mean difference of the “Yes” group and the “No” group reached significance for builder assumption scale, but was otherwise for the generated tutor behaviour scale. So, it can be said that the “Yes” group were more stringent than the “No” group in their reaction to both scales. Nevertheless, the views of the groups were still very positive towards the constructs measured.

Overall, while there were mean differences recorded across various group factors, the magnitude of these differences ranged between small (for factors that did not attain significance) to moderate for previous authoring tool experience factor (which reached significance in the builder assumption scale only). The conclusion that can therefore be made is that these factors did not influence the users’ position on constructs measured, except with respect to the previous experience with the authoring tools factor on builder assumption scale—although the patterns of the differences were consistent across the two scales measured. This implies that various groups held very similar views with respect to the objects evaluated. So, the views were not strictly biased by the independent variables or factors, although previous authoring tool experience counted moderately.

■Investigating the Relationship between BDASUM and TUTBHV Scales

Correlation analysis was employed to explore the existence of any relationship between the builder assumption scale and the generated tutor behaviour scales. This attempts to answer the third question with respect to the proposition under examination. However, due to lack of normality with respect to the scales under consideration (as discussed in the preliminary analysis—see appendix 5.1 for detail) and other stringent assumptions that should be fulfilled to use Pearson correlation, the Spearman correlation—a nonparametric statistic—was chosen. Although, as discussed in the preliminary analysis, the robustness of parametric statistics can accommodate non-normality, other conditions that need to be fulfilled to use Pearson correlation could not be ascertained. Also, Spearman correlation as been used and found reliable in many delicate research situations, such as medical, psychological and educational researches (Huson, 2007; Paternostro-Sluga et al., 2008; Longo et al., 2011). Hence, the research opted for the Spearman correlation to avoid any bias that the analysis may be subjected to, if the Pearson correlation is used. Thus, table 5.7 below presents an extract of the full result of the Spearman correlation analysis between builder assumptions scale/items and the generated tutor's behaviour scale/items.

Table 5. 7: The Spearman Correlation Analysis between BDASUM and TUTBHV scales

			Builder Assumptions scale	Behaviour scale
Spearman's rho	Builder Assumptions scale	Correlation Coefficient	1.000	.542**
		Sig. (2-tailed)	.	.000
		N	82	82
	Generated Tutor Behaviour scale	Correlation Coefficient	.542**	1.000
		Sig. (2-tailed)	.000	.
		N	82	82

**. Correlation is significant at the 0.01 level (2-tailed).

An inspection of the above table (table 5.7) and the full result (appendix 5.5) suggests:

- that correlation between the scales was statistically significant at 0.01 alpha level, rho=0.542 and p=0.0;
- that correlation between the scales items reached significance at 0.01 alpha level, except with scale item "bdasum3", which correlated with item "tutbhv3" at 0.05 alpha level, rho= 0.244, p=0.027;
- Correlations among items were all positive, with the high score of the independent variables (IVs) associated with a high score of the dependent variables (DVs);
- The strength of the relationship between the scales (builder assumption and generated tutor behaviour scales) was large (rho = 0.542—using Cohen, 1988, pp. 79–81, guidelines, which state as follows: small => r=0.10 to 0.29; medium => r=0.30 to 0.49; and large => r=0.50 to 1.0); thus, suggesting quite a strong positive relationship between builder assumption scale (IV) and generated tutor behaviour (DV).

Based on the above outcome, the coefficient of determination—an indicator of how much variance the two variables share—was calculated using the procedure recommended in Pallant (2010), thus:

$$\text{shared variance} = \text{rho}^2 \times 100\% = 0.542^2 \times 100\% = 29.38\ \%\ \text{approx.}$$

The computed shared variance discloses that the builder assumptions helped explain 29% of the variance of the generated tutor's behaviour. The shared variance achieved can be said to be a quite respectable amount of variance explained, compared with much of the research conducted in the social sciences, as noted in Pallant (2007). Also, the achieved result could not be corroborated or compared with that of educational related researches, since the level of explained variance was not known; this was therefore targeted in future researches.

In summary, it could be said that there is a statistically significant and strong positive correlation between the two variables, builder assumptions and generated tutor behaviour scales (rho = 0.542, n = 82, $p < 0.01$), and between the scales items (at $p < 0.01$). The only exception was the case involving one independent variable (scale item "bdasum3") and one dependent variable (scale item "TUTBHV3"); they achieved significant correlation at 0.05 alpha level. Also, the independent variable explained 29% of the variance in the dependent variable. In addition, high scores of builder assumptions scale/items explained the high scores recorded in the generated tutor behaviour scale/items.

5.3.1.2 Proposition 1.2

It is possible to generate tutoring systems that support process monitoring and model-tracing from the implemented metamodel (i.e. ILABS)

For the proposition, a confirmatory investigation into the tutoring strategies implemented in ILATs generated from ILABS was targeted. "TUTSTG" (tutoring strategies) scale addresses this aspect of the research and the following analyses aim to present users' reactions to this.

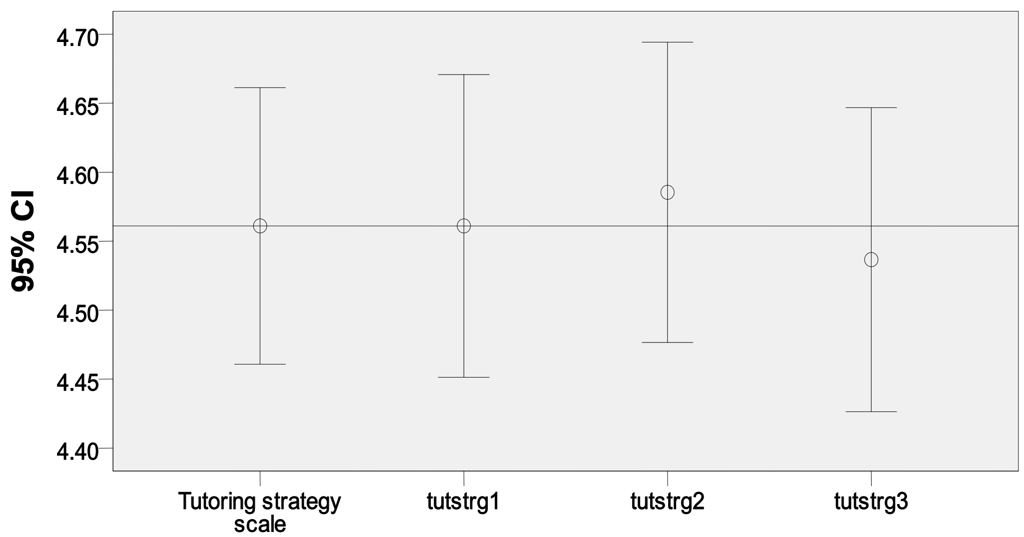
■Investigating Users Reaction/Perception—Univariate Analysis of Scale/Scale Items

Table 5.8 below provides the results of the univariate analysis of the scale in question, extracted from table 5.3 above. Users' reactions at the scale level record an overall reaction mean score of 4.56 and standard deviation of 0.456. The mean score was above the mid-score of a five-point Likert scale (the benchmark used). Likewise, mean scores for the scale items were above the benchmark, where item "tutstrg1"—*builder allows me to generate eTutors that monitors student's problem-solving steps (signposting cognitive mapping)*—mean=4.56, median=5.0, std.=0.499, min.=4, max.=5; item "tutstrg2"—*builder allows me to produce eTutors that traces student's input value (signposting model tracing)*—mean=4.59, std.=0.496, min.=4, max.=5; and item "tutstrg3"—*builder allows me to produce eTutors that monitor problem-solving steps* *and trace students input value*—mean=4.54, std.=0.502, min.=4, max.=5. It was observed that item "tutstrg2"—signposting model-tracing—had the mean score (4.59) greater than the scale mean overall reaction (4.56), and others were lower (or less). Nevertheless, scale item mean scores tend towards the same direction as the mean overall reaction of the scale, suggesting strong affirmative views across the scale items. It thus implies the builder implemented the constructs measured. However, can it be concluded that the mean overall reaction score of the scale and the mean scores of scale items are statistically different from the benchmark in order to conclude that users favour the construct measured? Secondly, can it be affirmed that that the mean scores of the scale items are not statistically different from the mean overall reaction score of the scale? Although the latter two points may be implied by the results presented in the table, further statistical investigation may be required to ascertain them.

Table 5. 8: Statistics of Users' Reaction to Builder Tutoring Strategy

Scale / Scale Items	N	Minimum	Maximum	Mean	Median	Std. Deviation
Tutoring Strategy scale	**82**	**4.00**	**5.00**	**4.56098**	**4.66667**	**.456257**
tutstrg1	82	4	5	4.56	5.00	.499
tutstrg2	82	4	5	4.59	5.00	.496
tutstrg3	82	4	5	4.54	5.00	.502

In response to the questions raised above, a simple error bar chart (see figure 5.3) seems to suggest: with respect to the second question—that the views are not statistically different from the overall reaction. On the other hand, the views were statistically different from the benchmark point (which is, neither agree nor disagree). Further statistical steps were taken to ascertain this claim, using one–sample t-test, as is discussed next.



■ **Analysis of Difference between Mean Scores and Benchmark Value**

In order to address the first question raised above, the extent of difference between mean scores and the benchmark was investigated via one-sample t-test, as shown in table 5.9 below. The results reveal that the mean score of the overall reaction to the scale was statistically significant, t (81)=30.981, p=0.0, and the magnitude of the difference (mean difference=1.56, 95% CI: 1.46 to 1.66) was very large (eta squared=0.92). Likewise, all the mean scores of the scale items reached statistical significance. Thus, it suggests that means were largely different from the benchmark (3.0).

Table 5. 9: One Sample t-test—Comparison of Mean Scores and Benchmark

	Test Value = 3.0 (benchmark)					
					95% Confidence Interval of the Difference	
	T	Df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Tutoring strategy scale	30.981	81	.000	1.560976	1.46073	1.66123
tutstrg1	28.309	81	.000	1.561	1.45	1.67
tutstrg2	28.962	81	.000	1.585	1.48	1.69
tutstrg3	27.733	81	.000	1.537	1.43	1.65

■ **Analysis of Difference between Scale and Scale Item Mean Scores**

Similar to the above, the one-sample t-test was employed, yet unlike the above, the mean overall reaction score of the scale was adopted as the test value (instead of the benchmark). Table 5.10 shows the outcome of the test. The result at 0.05 alpha level, therefore, indicates no significant statistical difference between the scale mean score and the scale items mean scores: for item “tutstrg1”—t (81)=0, p=1, and there was no difference (mean difference=0, 95% CI: -0.11 to 0.11); for item “tutstrg2”—t (81)=0.445, p=0.657, and the magnitude of the difference (mean difference=0.024, 95% CI: -0.08 to 0.13) was very small (eta squared=0.002); and for item “tutstrg3”—t (81)=-0.440, p=0.661, and the magnitude of the difference (mean difference=-0.024, 95% CI: -0.13 to 0.09) was very small (eta squared=0.002). Thus, the results suggest that the views expressed at the scale level were a replica of the scale items.

Since no significant difference exists between the scale items mean scores and scale overall reaction mean, but there exists a significant difference between the overall mean score and the benchmark, it can therefore be concluded that the overall response thus tends towards ‘strongly agree’. This signifies that users were favourably disposed to the claim that ILABS implemented the constructs measured. Hence, ILATs that implement the tutoring strategies—process monitoring and model tracing strategies—can be generated.

Table 5. 10: One-Sample Test—Comparison of Scale and Scale Item Means

	Test Value = 4.56098					
					95% Confidence Interval of the Difference	
	T	Df	Sig. (2-tailed)	Mean Difference	Lower	Upper
tutstrg1	.000	81	1.000	.000	-.11	.11
tutstrg2	.445	81	.657	.024	-.08	.13
tutstrg3	-.440	81	.661	-.024	-.13	.09

5.3.2 Reaction Evaluation of Key Builder Features

Within this category, only one proposition is considered. It addresses three constructs, namely flexibility to create multiple tutors for numerical topics and their variants, prior skills required to use the ILABS effectively, and likely production time. Although these attributes can be examined under the usability scale (treated in proposition 1.4), the constructs were singled-out and studied via separate scales, to gain deeper insight, and to address aspects of research question one related to them. Details of the scale items that constitute the constructs are provided in table 5.11. Analyses of these constructs are rendered in the following sub-sections.

Table 5. 11: Description of Scale Items

Builder Restriction scale (BDRST):		
1	bdrst1	The builder system allows me to generate as many eTutors as I wish.
2	bdrst2	The builder system allows me to produce eTutors for different topics (or modules).
3	bdrst3	The builder system allows me to generate variants of an eTutor
Production Time scale (PDTIM):		
1	pdtim1	The builder system allows me to configure and generate an eTutor within a short span of time.
2	pdtim2	The builder system takes more than five hours to configure and generate an eTutor for a one-hour tutorial session.
Special Skills scale (SPSKL):		
1	spskl1	I need computer programming skill to be able to generate an eTutor from the builder system.
2	spskl2	Knowledge of accounting is required to generate meaningful eTutors from the builder system.

5.3.2.1 Proposition 1.3

The implemented metamodel can be used to create an unrestricted number of tutoring systems within a short space of time by non-programmers

As shown in table 5.11 above, this proposition addresses three constructs via the following builder questionnaire scales: “BDRST”—builder restriction scale,“PDTIM”—production time scale, and “SPSKL”—special skills scale. In order to determine users' reactions to them, various analyses were employed, and each scale was discussed separately. Table 5.12 below presents results of univariate analysis: mean score, standard deviation, minimum and maximum scores of users’ reactions to the prototype ILABS with respect to constructs under consideration. The overall mean scores for scales, as well as their items' mean scores were included in the table.

[A] For Builder Restriction scale (BDRST):

The mean overall reaction score for the scale was greater than the benchmark (M=4.38, SD=0.581, min.= 3.0, max.= 5.0); likewise, the mean scores of the scale items: item “bdrst1”—‘*builder allows me to generate as many eTutors as I wish*’ (M=4.37, SD=0.639, min.=2 and max.=5); item “bdrst2”—‘*builder allows me to produce eTutors for different topics or modules*’ (M=4.60, SD=0.593, min.=3 and max.=5); and item "bdrst3"—'*builder allows me to generate variants of an eTutor*' (M=4.26, SD=0.750, min.=3 and max.=5).

Table 5. 12: Descriptive Statistics for BDRST, PDTIM & SPSKL scales

	N	Minimum	Maximum	Mean	Std. Deviation
Builder Restriction scale	**82**	**3.000**	**5.000**	**4.38211**	**.581191**
bdrst1	82	2	5	4.37	.639
bdrst2	82	3	5	4.52	.593
bdrst3	82	3	5	4.26	.750
Production Time scale	**82**	**2.000**	**5.000**	**3.84756**	**.796168**
pdtim1	82	2	5	4.26	.625
pdtim2	82	1	5	3.44	1.218
Special Skill scale	**82**	**2.500**	**5.000**	**4.28049**	**.648348**
spskl1rvs	82	1	5	4.09	1.009
spskl2	82	3	5	4.48	.571
Valid N (listwise)	82				

Although, the above results suggest an affirmation of the construct measured, a close look into the table shows that the mean score for item "bdrst2" was greater than the mean overall reaction score of the scale; the other two were lower than the scale mean score. However, whether they are significantly different from the mean overall reaction score of the scale needs further investigation to determine the extent of the difference—if any, and whether they all represent the same view. Equally, it was necessary to determine the existence of any significant difference between the mean scores and the benchmark, in order to accept this aspect of the proposition. As a result, one-sample t-test was utilised to investigate the aforementioned issues. The results are thus presented in tables 5.13 and 5.14 below.

■Analysis of Difference between Benchmark Value and Mean Scores

From table 5.13, there was statistical difference between the mean scores of the scale and the benchmark, t (81)=21.534, p=0; the magnitude of the difference (mean difference=1.38, 95% CI: 1.25 to 1.51) was very large (eta squared= 0.85). In the same light, all the scale items attained statistical significance; their p-values equal zero (0). Therefore, the results further align with the earlier stated implication, i.e. affirmation of the constructs measured.

Table 5. 13: One-Sample Test—Comparison of Mean Scores and Benchmark value

	Test Value = 3.0 (benchmark)					
					95% Confidence Interval of the Difference	
	T	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Builder Restriction scale	**21.534**	**81**	**.000**	**1.382114**	**1.25441**	**1.50982**
bdrst1	19.370	81	.000	1.366	1.23	1.51
bdrst2	23.291	81	.000	1.524	1.39	1.65
bdrst3	15.156	81	.000	1.256	1.09	1.42

■Analysis of Difference between Scale and Scale Items Mean Scores

Further to the above, one-sample t-test was employed to examine the existence of difference between the mean score of the builder restriction scale and mean scores of its items. The results from table 5.14 show that there was no significant difference between the mean score of the scale and two of the scale items at 0.05 alpha level, thus: item "bdrst1"—t (81)= -0.231, p=0.818, and the magnitude of the difference (mean difference=-0.016, 95% CI: -0.16 to 0.12) was extremely small to be noticed (eta squared=0.0007); item "bdrst3"—t (81)=-1.520, p=0.132, and the magnitude of the difference (mean difference=-0.126, 95% CI: -0.29 to 0.04) was small (eta squared=0.03).

Table 5. 14: One-Sample Test—Comparison of Scale and Scale Items Mean Scores

	Test Value = 4.38211 (Mean Score of Scale)					
					95% Confidence Interval of the Difference	
	T	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
bdrst1	-.231	81	.818	-.016	-.16	.12
bdrst2	2.174	81	.033	.142	.01	.27
bdrst3	-1.520	81	.132	-.126	-.29	.04

On the other hand, statistical significance was attained at 0.05 alpha level with respect to item "bdrst2", t (81)=2.174, p=0.033; however, the magnitude of the difference (mean difference=0.142, 95% CI: 0.01 to 0.27) was moderate (eta squared=0.06). Despite the difference, it does not in any way refute the ILABS capability to produce ILATs for different topics. It establishes existence of strong views towards the latter item (bdrst2), compared to other two items ("bdrst1" and "bdrst3") in the scale. Collectively, therefore, it can be said that the three items affirm the construct measured by the scale.

From the above analysis, users' views can be said to be similar across the board, since no large differences occur between items and the scale; that these views tend around the overall reaction. These views were very strong in favour of the construct measured. Hence, this part of the proposition under consideration—referring to the builder restriction construct in proposition 1.3—is thereby accepted.

[B] Production Time scale (PDTIM):

This had an overall mean reaction score of M=3.85—which is above the benchmark, standard deviation, SD= 0.80, minimum score, min.=2.0 and maximum score, max.=5.0. Statistics for the scale items indicate as follows: item pdtim1—*builder allows me to configure and generate an eTutor within a short span of time* (M=4.26, SD=0.625, min.=2 and max.= 5); item "pdtim2"—*builder allows me to configure and generate in less than five (5) hours an eTutor meant for one-hour of tutorial session* (M=3.44, SD= 1.218, min.=1 and max.=5). Similar to the scale, both items' mean score were above the benchmark. However, the mean score of item "pdtim1" (M=4.26) seems greater than that of the scale (3.85) and item "pdtim2" (3.44), while item "pdtim2" had the lowest, thereby suggesting an investigation into the significance of the differences and their individual impact on the mean overall reaction score of the scale. Apart from scale item "pdtim1" with mean score above 4.0, the other scale item—"pdtim2"—and mean overall reaction score of the scale were both close to the threshold (3.0). The significant difference between the latter means and the threshold also needs to be determined, in order to have a solid platform for the conclusion that might be reached.

■Analysis of Difference between Means and Benchmark

From table 5.15, the one-sample test indicated statistical significance between the benchmark and mean overall reaction of the production time scale at 0.05 alpha level, t (81)=9.640, p=0; the magnitude of the difference (mean difference=0.848, 95% CI: 0.673 to 1.023) was large (eta squared=0.53). Equally, the scale items reached statistical significance, item "pdtim1"—t (81)=18.204, p=0, and the magnitude of the difference (mean difference=1.256, 95% CI: 1.12 to 1.39) was very large (eta squared=0.80); and item "pdtim2"—t (81)=3.264, p=0.002, and the magnitude of the difference (mean difference=0.439, 95% CI: 0.17 to 0.71) was moderate (eta squared=0.12). From this result, it shows that users' reactions to the scale items were not balanced. Reaction to the first item was extremely strong, indicating that users believe the eTutor can be generated within a short span of time. On the other hand, they seem not to strongly agree that eTutor can be generated within five hours, though there was meaningful mean difference, which suggests affirmation of the issue raised in the item. Nevertheless, overall reaction indicates that the ILABS is supportive of the construct measured, but efforts must be made to improve this aspect of the builder.

Table 5. 15: One-Sample Test—Comparison of Mean Scores and Benchmark value

	Test Value = 3.0 (benchmark)					
					95% Confidence Interval of the Difference	
	T	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Production Time scale	**9.640**	**81**	**.000**	**.847561**	**.67262**	**1.02250**
pdtim1	18.204	81	.000	1.256	1.12	1.39
pdtim2	3.264	81	.002	.439	.17	.71

■Analysis of Difference between Scale and Scale Items Mean Scores

The above analysis, clearly tells the story of moderate to large differences between the scale/items and the benchmark. Existence of statistical difference between the scale and items is not known, although the above result may suggest one. Investigation of the latter was conducted using a one-sample t-test, with mean score of the scale used as the test value. The results, as shown in table 5.16 below, signify statistical significance at 0.05 alpha level, between the mean scores of the scale items and the scale, thus: item "pdtim1"—t (81)=5.921, ρ=0, and the magnitude of the difference (mean difference=0.409, 95% CI: 0.27 to 0.55) was large (eta squared=0.40); and item "pdtimt2"—t (81) = -3.037, ρ = 0.003, and the magnitude of the difference (mean difference=—0.409, 95% CI:—0.68 to—0.14) was moderate (eta squared=0.10). This result further confirmed the earlier stated analysis, since users' reactions to both items were at different levels.

Table 5. 16: One-Sample Test: Comparison between Mean Scores of Scale Items and Scale

	Test Value = 3.84756					
					95% Confidence Interval of the Difference	
	T	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
pdtim1	5.921	81	.000	.409	.27	.55
pdtim2	-3.037	81	.003	-.409	-.68	-.14

Conclusively, it was observed that users' views were widely apart with respect to the scale items. However, these views were positive. While users agree that ILABS allows them to configure/generate an eTutor (i.e. ILAT) within a short span of time, many were sceptical about ILABS allowing them to configure/generate an ILAT in less than five hours for a tutorial session that lasts for one (1) hour. An inspection of the relevant frequency distribution further explains the pattern of the scepticism with respect to the latter item (item “pdtim2”)—see appendix 5.6; it shows that only 52.4% respondents express agreement (26.8%—agree and 25.6%—strongly agree). While others, mainly disagree (31.7%) or neither agree nor disagree (14.6%). Despite the pattern, it can still be concluded that ILABS enhances the configuration and generation of tutoring systems, but the turnaround time needs to be revisited and refined/improved, more so since ILABS is still a prototype.

***[C]* Special Skills scale (SPSKL):**

This is a two-item scale; its descriptive statistics indicate that M=4.28, SD=0.65, min.=2.5 and max.=5.0. The scale items statistics for items “spskl1rvs”—*I need computer programming skills to be able to generate an eTutor from the builder system* are (M=4.09, SD=1.01, min.=1 and max.=5); and for item “spskl2”—*knowledge of accounting is required to generate meaningful eTutors from the builder system* (M=4.48, SD=0.571, min.=3 and max.=5). The mean scores reported were all above 4.0; suggesting general agreement on the constructs measured on a five-point Likert scale. Although they differ in magnitude, the effect of the difference is not known. Further investigation was conducted to reveal existence of any significant differences and their impact on mean overall reaction score of the scale. Also, an investigation into the differences between the means and the benchmark was launched; it aimed at confirming or refuting the earlier claim that the descriptive statistics suggested—i.e. agreement of users on construct measured.

■Analysis of Difference between Benchmark and Mean Scores

One-sample t-test was employed to compare the mean scores of the scale/items and benchmark. The result, see table 5.17 below, revealed that a statistically significant difference was attained between the mean score of the overall reaction to the scale and the benchmark at 0.05 alpha level, t (81)=17.884, p=0, and the magnitude of the difference (mean difference=1.28, 95% CI: 1.138 to 1.423) was very large (eta squared=0.798). Equally, the scale items reached statistical significance, item “spskl1rvs”—t (81)=9.744, p=0, the magnitude of the difference (mean difference=1.085, 95% CI: 0.86 to 1.31) was very large (eta squared=0.540); and item “spskl2”—t (81)=23.383, p=0, and the magnitude of the difference (mean difference=1.476, 95% CI: 1.35 to 1.60) was very large (eta squared=0.871). Due to the large effect size, these results suggest strong affirmation of the constructs measured by the scale.

Table 5. 17: One-Sample Test—Comparison of Mean Scores and Benchmark value

	Test Value = 3.0 (benchmark)					
					95% Confidence Interval of the Difference	
	T	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Special Skill scale	**17.884**	**81**	**.000**	**1.280488**	**1.13803**	**1.42295**
spskl1rvs	9.744	81	.000	1.085	.86	1.31
spskl2	23.383	81	.000	1.476	1.35	1.60

■Analysis of Difference between Scale Mean and Scale Items Mean Scores

An investigation of the differences between mean score of the scale and that of the scale items (see table 5.18 below), shows no statistical difference between the mean score of the overall reaction to the special skills scale and the scale item "spskl1rvs" at 0.05 alpha level: t (81)= -1.752, ρ=0.084, and the magnitude of the difference (mean difference= -0.195, 95% CI: -0.42 to 0.03) was small (eta squared=0.037). On the other hand, there was statistical significance between the mean score of the overall reaction to the scale and item "spskl2" at 0.05 alpha level: t (81)=3.092, ρ=0.003; however, the magnitude of the difference (mean difference=0.195 at 95% CI: 0.07 to 0.32) was moderate (eta squared=0.106). Accordingly, while the effect of mean score for item "spskl1rvs" was small and it explained only 3.7% of the variance in mean overall reaction, the effect of item "spskl2" was moderate and it explained 10.6% of the variance in the mean overall reaction—the basis of comparison.

Table 5. 18: One-Sample Test—Comparison between Mean Scores of Scale and Scale Items

	Test Value = 4.28049 (mean overall reaction score of scale)					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
spskl1rvs	-1.752	81	.084	-.195	-.42	.03
spskl2	3.092	81	.003	.195	.07	.32

Although there was general acceptability for the skills constructs measured in the scale, there was significant difference between the overall reaction to the scale and reaction to the need for accounting skill. Thus, it indicates that the overall reaction, more-or-less, represents the views of users that 'prior knowledge of accounting skills is required' to use the ILABS. Also, the need for programming skills, as a condition to use ILABS was rejected.

5.3.3 Usability Evaluation

The last proposition within research question one deals with the evaluation of ILABS, from the perspective of ease of use and general usability of the prototype system. It intends to determine the perception of users on the usability construct. The following gives details of the analysis carried out.

5.3.3.1 Proposition 1.4

> Users of implemented metamodel have a positive perception about its ease of use and usability

This proposition is examined through the usability scale, “USAB”, of the builder questionnaire. It contains nineteen items, but eight of the items were negatively worded. During coding, scores for negatively worded items were reverse-coded (Pallant, 2010), to enhance application of relevant statistics. Accordingly, users' responses to negatively worded items, scores one (1), two (2), three (3), four (4), and five (5) were re-coded in reverse order thus: five (5), four (4), three (3), two (2), and one (1) respectively. Thereafter, relevant statistics were applied to the coded data, as discussed below.

■Investigating Users’ Reaction/Perception—Univariate Analysis/One-Sample Test

In order to examine the usability issues, univariate statistics and parametric statistics were employed at different instances based on certain considerations discussed subsequently. The univariate statistics revealed mean (3.96) and median (3.89) of the scale (see appendix 5.7). Both values were greater than the benchmark (3.0), suggesting agreement on the usability construct under examination, unless proven otherwise. The values were compared via a one-sample t-test, to determine the extent of significance in difference, if any, and to aid subsequent analysis. The result of the test, table 5.19 below, shows that there is no statistically significant difference between the two statistical measures (i.e. the mean and median) at 0.05 alpha level, t (81)=1.764, p=0.081 and magnitude of the difference (mean difference=0.069, 95% CI:-0.009 to 0.147) was small (eta squared=0.04). The result thus paves the way for either measure to be used in subsequent tests conducted to examine users' views. As a result of the revelation, mean measure was chosen as a basis for comparison in subsequent analysis; this enabled the determination of the strength of users’ reaction to the usability construct of ILABS.

Scale items level analysis, using the mean measure, revealed that twelve (12) items had mean scores above 4.0, while seven (7) items were below (see table 5.20 below). Out of thesse seven (7), only one had its mean score below the benchmark 3.0, which ordinarily signposts disagreement. However, since the latter item in question is negatively worded, its mean score was reverse-interpreted; this applies to other negatively-worded items. For example, if the mean score signifies agreement, it will be interpreted as disagreement, and vice-versa. Hence, it can be said that the item in question signposts agreement on the usability issue treated.

Table 5. 19: One-Sample t-Test—Comparison of Mean and Median Scores of Usability Scale

	Test Value = 3.89474 (Median score)					
					95% Confidence Interval of the Difference	
	T	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Usability scale	1.764	81	.081	.069316	-.00886	.14749

Furthermore, a one-sample t-test conducted to compare the mean scores of the usability scale/items and the benchmark, indicated a statistically significant difference at 0.05 alpha level (see table 5.21 below); the t-test result for the usability scale is given by $t(81)=24.536$, $p=0$, and the magnitude of the difference (mean difference=0.964, 95% CI: 0.886 to 1.042) was very large (eta squared=0.88). Based on the above, it can be concluded that users—on a general level—agree to the usability construct, i.e. ILABS is usable having satisfied the necessary ease of use and usability conditions.

On a per scale item basis, it could also be concluded that users agreed to all the eleven positively-worded items (usab4, usab5, usab7, usab9, usab10, usab11, usab13, usab15, usab17, usab18, and usab19); while they disagreed with the negatively worded items (usab1, usab2, usab3, usab6, usab8, usab14, usab16), with the exception of item "usab12"—"*system needs more introductory explanation*", M=2.07, SD=0.716; $t(81)=-11.717$, $p=0.0$, and the magnitude of the difference (mean difference=-0.927, 95% CI: -1.08 to -0.77) was large (eta squared=0.63). The implication, therefore, is that users opined that introductory support will be required to use the system. The explanation that can be provided, is that more work needs to be done to fine-tune ILABS, being the first version. Such work should include provision of help facilities that can reduce human expert intervention/support that users may require.

Table 5. 20: One-Sample Statistics for USAB scale

	N	Mean	Std. Deviation	Std. Error Mean
Usability scale	**82**	**3.96406**	**.355802**	**.039292**
usab1rvs - Builder can be described as annoying	82	4.27	.668	.074
usab2rvs - Builder can be described as confusing	82	4.04	.693	.077
usab3rvs - Builder can be described as frustrating	82	4.11	.667	.074
usab4 - Builder can be described as interesting	82	4.35	.575	.063
usab5 - Builder can be described as stimulating	82	4.37	.533	.059
usab6rvs - Builder can be described as tiresome	82	3.76	.695	.077
usab7 - Builder cab be described as usable	82	4.37	.639	.071
usab8rvs - Builder can be described as unpleasant	82	4.06	.635	.070
usab9 - I feel in control when I am using the system	82	3.96	.728	.080
usab10 - Builder system uses terms that are understandable	82	4.01	.923	.102
usab11 - Builder system uses terms that are familiar to me	82	3.63	.794	.088
usab12rvs - Builder system needs more introductory explanations	82	2.07	.716	.079
usab13 - It is easy to understand the objects on the Builder system's interface	82	4.15	.569	.063
usab14rvs - Builder system is slow	82	3.79	.698	.077
usab15 - I get what I expect when I click on objects on the Builder system interface	82	3.90	.730	.081
usab16rvs - It is difficult to move around the Builder system	82	3.28	.959	.106
usab17 - I feel efficient when using the Builder system	82	4.16	.555	.061
usab18 - Builder system can be characterised as innovative	82	4.55	.570	.063
usab19 - Overall, I am satisfied with the Builder system	82	4.49	.593	.065

Table 5. 21: One-Sample Test – Comparison of USAB Scale/Items Means Scores and Benchmark

	Test Value = 3.0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Usability scale	24.536	81	.000	.964056	.88588	1.04223
usab1rvs - Builder can be described as annoying	17.204	81	.000	1.268	1.12	1.41
usab2rvs - Builder can be described as confusing	13.547	81	.000	1.037	.88	1.19
usab3rvs - Builder can be described as frustrating	15.071	81	.000	1.110	.96	1.26
usab4 - Builder can be described as interesting	21.333	81	.000	1.354	1.23	1.48
usab5 - Builder can be described as stimulating	23.198	81	.000	1.366	1.25	1.48
usab6rvs - Builder can be described as tiresome	9.852	81	.000	.756	.60	.91
usab7 - Builder cab be described as usable	19.370	81	.000	1.366	1.23	1.51
usab8rvs - Builder can be described as unpleasant	15.122	81	.000	1.061	.92	1.20
usab9 - I feel in control when I am using the system	11.989	81	.000	.963	.80	1.12
usab10 - Builder system uses terms that are understandable	9.932	81	.000	1.012	.81	1.21
usab11 - Builder system uses terms that are familiar to me	7.235	81	.000	.634	.46	.81
usab12rvs - Builder system needs more introductory explanations	-11.717	81	.000	-.927	-1.08	-.77
usab13 - It is easy to understand the objects on the Builder system's interface	18.237	81	.000	1.146	1.02	1.27
usab14rvs - Builder system is slow	10.282	81	.000	.793	.64	.95
usab15 - I get what I expect when I click on objects on the Builder system interface	11.187	81	.000	.902	.74	1.06
usab16rvs - It is difficult to move around the Builder system	2.648	81	.010	.280	.07	.49
usab17 - I feel efficient when using the Builder system	18.907	81	.000	1.159	1.04	1.28
usab18 - Builder system can be characterised as innovative	24.611	81	.000	1.549	1.42	1.67
usab19 - Overall, I am satisfied with the Builder system	22.718	81	.000	1.488	1.36	1.62

■Investigating Main Effect of Demographic/Computer Experience Factors

The analysis above provides a general outlook to the perception of users on usability/ease of use constructs measured. Further investigation, to determine the effect of demographic/computer experience factors on users' reaction was carried out using t-test and one-way ANOVA as appropriate. It thus helps to determine if users' reactions were biased by any of the factors. Factors considered included institutions, department, gender, and computer experience—specifically, authoring tool experience. Instance(s) where significant difference(s) were recorded, the extent/impact of the differences on overall reaction mean of the scale were equally determined.

***[a]* Effect of Institution & Department Factors**

The effects of institutional and department factors were investigated using one-way ANOVA, since there were three groups per factor. Lavene's test of homogeneity of variance was not significant for both factors, institution (F=2.991, Sig.=0.056) and department (F=2.034, Sig.=0.138). Thus, the assumption of homogeneity of variance was not violated in both factors, giving credence to the applicability of one-way analysis of variance (ANOVA) in this section of the work. With respect to institution factor, table 5.22 revealed a statistically-significant main effect at 0.05 alpha level, $F(2, 81)=5.436$, $p=0.006$; the actual difference in mean scores between the groups was moderate (the effect size calculated, using eta squared, was 0.12) using Cohen's (1988, pp. 248-7) terms.

Table 5. 22: One-way ANOVA—Effect of Institution on Usability scale

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	1.240	2	.620	5.436	.006
Within Groups	9.014	79	.114		
Total	10.254	81			

On the other hand, the effect of department on usability scale (table 5.23) did not reach statistical significance, $F(2, 81)=0.109$, $p=0.897$; the actual difference in mean scores between groups was very small (the computed effect size, using eta squared, is 0.003).

Table 5. 23: One-way ANOVA—Effect of Department on Usability scale

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	.028	2	.014	.109	.897
Within Groups	10.226	79	.129		
Total	10.254	81			

The above results show that users' opinions or perceptions across institutions differ, while no significant differences occur across departments. While institution explained 12% of the variance in usability scale overall reaction, department could only explain 0.3% of the variance. What this implies, therefore, is that the department or discipline that users belong to, did not count in their overall reaction to the ILABS' usability construct, but the institution did count.

Further investigation, aimed at locating the institution(s) that is/are having a significant effect on overall reaction to the usability scale was conducted by inspecting the post-hoc test result (see table 5.24). The post-hoc test using the Tukey HSD test revealed that the mean score of institution Uni.B (M=3.82, SD=0.267) was significantly different from institution Uni.D, (M=4.11,SD=0.372), p.=0.006 at 0.05 alpha level, and the mean difference=±0.288; but institution Uni.C (M=4.01, SD=0.381) did not differ significantly from either institution Uni.B (p=0.088, mean difference=±0.193) or Uni.D (p=0.584, mean difference=±0.095).

Table 5. 24: Multiple Comparisons using Tukey HSD test—Main Effect of Institutions on Usability

(I) Institution	(J) Institution	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
2 Uni.B	3 Uni.C	-.193092	.090164	.088	-.40846	.02228
	4 Uni.D	-.287829*	.090164	.006	-.50320	-.07246
3 Uni.C	2 Uni.B	.193092	.090164	.088	-.02228	.40846
	4 Uni.D	-.094737	.095540	.584	-.32295	.13348
4 Uni.D	2 Uni.B	.287829*	.090164	.006	.07246	.50320
	3 Uni.C	.094737	.095540	.584	-.13348	.32295

*. The mean difference is significant at the 0.05 level.

Although the three institutions had mean scores above the benchmark (3.0), an indication of positive reaction to the usability of the builder system, the descriptive statistics (see table 5.25 below) revealed that institution Uni.B had the lowest mean (M=3.82, SD=0.267) while institution Uni.D had the highest mean (M=4.11, SD=0.372). Both values impacted the overall mean reaction of the scale; Uni.B had a downward effect, while Uni.D had an upward effect on the scale's mean score. These results therefore inform that users views, though all were positive with respect to the usability constructs, they were lowest in institution Uni.B and highest/strongest in institution Uni.D.

Table 5. 25: Descriptive Statistics of Usability Scale by Institutions

Institution		N	Minimum	Maximum	Mean	Std. Deviation
2 Uni.B	Usability scale	32	3.316	4.368	3.81743	.267363
	Valid N (listwise)	32				
3 Uni.C	Usability scale	25	3.263	4.579	4.01053	.380746
	Valid N (listwise)	25				
4 Uni.D	Usability scale	25	3.158	4.579	4.10526	.371851
	Valid N (listwise)	25				

[b] Effect of Gender Factor

The effect of gender on users' reaction to usability scale was investigated using independent samples t-test (see appendix 5.8). Lavene's test of equality of variances indicated no significance, F=1.370, Sig.=0.245, thereby confirming the applicability of independent samples t-test in this context. The result of the test between the mean scores of male users (M=4.00, SD=0.324) and female users (M=3.86, SD=0.411), indicated no significant difference at 0.05 alpha level, thus: t (80)=1.730, p=0.087; the magnitude of the differences (mean difference=0.148, 95% CI: -0.022 to 0.317) was small (eta squared=0.04) from the perspective of Cohen's (1988) criterion. Hence, it can be concluded that users' views align across gender groups.

[c] Effect of Computer Experience Factor

On the effect of authoring tool experience, **i**ndependent samples t-test was equally applied (see appendix 5.9). Lavene's test of equality of variances indicated no significance (F=0.462, Sig.=0.499), thus confirming the applicability of the t-test. The outcome of the t-test between the means of those with authoring experience, the "Yes" group (M=4.00 SD=0.369) and those without experience, the "No" group (M=3.93 SD=0.340), suggested no significant differences at 0.05 alpha level, t (79)=0.943, p=0.349; the magnitude of the difference (mean difference=0.075 approximately 95% CI:-0.083 to 0.233) was small (eta squared=0.01), based on Cohen's (1988) effect size criterion. As a result, it can be stated that presence/absence of authoring experience did not have any positive/negative impact on users' positions on the usability of the ILABS.

In summary, the above usability analyses suggest that:

- users' views were generally positive on the usability constructs measured;
- none of the factors influenced the opinion of users about the ILABS, except the institutional variable with respect to two of the institutions that participated. Reasons for this difference may require further investigation, which future research should address.

5.4 Qualitative Analysis

This section presents analysis of the qualitative data collected in the course of this research. It involves analysis of interviews conducted for lecturers in accounting and allied schools/departments of some higher education institutions. This, therefore, requires qualitative analytical method(s), which can enhance the emergence of rich findings from the qualitative data. The outcomes would enhance those from quantitative data, in an attempt to provide answer(s) to research question one (stated earlier in the chapter).

In the light of the above, several qualitative analytical methods, such as thematic analysis, content analysis etc., were identified in the literature (Harwood & Garry, 2003; Hsieh & Shannon, 2005; Braun & Clarke, 2006; Cassell et al., 2006). However, this research employs thematic-content analysis, which is further discussed below. The methods had been, invariably, further classified under two fundamental approaches according to their application (Spencer, Ritchie & O'Connor, 2004; Fereday & Muir-Cochrane, 2006; Lathlean, 2006; Burnard et al., 2008; McMillan, 2009). For instance, Burnard et al. noted that *"there are two fundamental approaches to analysis of qualitative data (although each can be handled in a variety of different ways): the deductive approach and the inductive approach."* (Burnard et al., 2008, p. 429).

As Burnard et al. (ibid) asserted, deductive approaches entail the use of a predefined framework, structure or theoretical concepts to identify themes from a data set. The approach imposes a predefined structure on data, which is then used to analyse the interview transcripts (Williams, Bower, & Newton, 2004). It tends to introduce bias and/or threatens the emergence of theme(s) not captured in such a framework or theoretical concepts, and is therefore prone to losing sight of other themes that might have emerged or aid theory development (Fereday & Muir-Cochrane, 2006; Burnard et al., 2008). However, its usefulness has been traced to studies where the researcher is aware of probable participants' responses, but sought to explore reasons for such responses (ibid). In contrast, the inductive approach enhances natural emergence of themes (Fereday & Muir-Cochrane, 2006). By implication, no predefined framework, structure or theory is assumed or imposed on data. So, themes emerge in a naturalistic manner, which helps promote theory development. Hence, it is considered suitable for grounded theory research (Burnard et al., 2008).

The qualitative aspect of this research, though not a grounded theory research, employs thematic-content analysis of interview transcripts, thus using both deductive and inductive approaches. By implication, two-stage analysis was undertaken: stage one employed an inductive approach, using thematic analysis to evolve themes in a natural way; then stage two applied a deductive approach, screening emergent themes using predefined concepts from the research's pedagogic metamodel and the literature. Hence, the limitations of the deductive approach were addressed. Moreover, the approach is very relevant in a research that is underpinned by theory, such as this work. By combining both approaches, the research aims to benefit from their complementary roles, as evidenced in previous studies that adopted both approaches (Fereday & Muir-Cochrane, 2006; Burnard et al., 2008; McMillan, 2009).

Based on the above, thematic-content analysis of interview transcripts was undertaken, to enhance natural emergence of themes and their subsequent grouping according to theoretical concepts considered. This aimed to eliminate bias that might have resulted from the use of the deductive approach alone. Subsequent sub-sections present a discussion of the characteristics of the interviewees and the themes that emerged from the analysis of the data collected. As stated earlier, the themes are categorised and presented according to propositions, which they addressed. Thereafter, a detailed discussion of the findings from the qualitative and quantitative analysis is given, in relation to the metamodel that underpins current research, and in the light of previous research studies and their implications for research and practice.

5.4.1 The Analytical Process / Characteristics of Interviewees

This sub-section presents analysis of the responses to the interview questions. This was undertaken with respect to the prototype ILABS and the sample ILAT generated. It therefore draws on the views and perception of lecturers (target users of the ILABS being evaluated), aimed at gaining better understanding of what the tools represent in relation to the metamodel that underpins their development, and teaching and learning or pedagogy in the numerical problem-solving context of the accounting and finance discipline. In order to achieve the above, interviews were conducted. However, despite efforts to hold, at least, three (3) interviews in each of the four institutions and across accounting allied departments of their business schools, only eight lecturers volunteered to participate in the interview process that was audio-taped with their permission. The eight lecturers, formally exposed to the prototype ILABS and the sample ILAT generated from it, were mainly from the department of accounting of the two institutions that participated—namely: Uni. B & Uni. D.

The audio-taped interviews were transcribed and analysed, benefiting from previous works in the literature (see Miles & Huberman, 1994; Lacey & Luff, 2001; Fer, 2004; Braun & Clarke, 2006; Auld et al., 2007; Burnard et al., 2008; Costu, Aydin, & Filiz, 2009; McMillan, 2009). The transcripts were read through several times while listening to the audiotapes to ensure accuracy. Thereafter, each transcript was treated by identifying core data bits from raw text using a word processor, Microsoft Word; these data bits constituted the level one codes, from which subsequent and higher level codes emerged. After the first or initial coding in Microsoft Word, the eight transcripts were merged and imported into Microsoft Excel for further coding that yielded higher layers of codes, i.e. levels two to four codes. Each code in an upper layer involves categorising or grouping together similar lower layer codes, under the same name or phrase. Thus, though the process involved many iterations, a better understanding of the issues being investigated was achieved and the final categories that emerged were related back to the propositions which they intended to address. It also affords views from different participants on the same issue to be compared, in order to identify similarities and differences among their perceptions and/or evaluation.

Table 5. 26: Demographic Characteristics & Computer Experience of Interviewees

Characteristic		**N**	**%**
Gender	Male	6	
	Female	2	
Work place	Uni.B	5	
	Uni.D	3	
Work Department	Accounting	8	100
	Other Depts.	---	---
Job Role	Lecturers	8	100
	Others	---	---
Highest Qualification	Professional	---	---
	Postgraduate	8	100
General Computer Experience	Yes	8	100
	No	---	100
Previous Authoring Tool Experience	Yes	---	---
	No	8	100
Previous e-Tutoring Experience	Yes		
	No		

From the analysis, six males and two females participated, and they all held qualifications to the level of postgraduate degrees, as well as having relevant professional qualifications (see table 5.26 above). Each of them had varied experiences, both in academia and in the industry, in the line of accounting, which is the domain utilised for evaluation of the tools developed in this research. In terms of computer experience, all the eight lecturers were computer literate; none of them had used any authoring tool before their exposure to the current tool being evaluated. However, some had used e-tutoring software before this research was introduced to them.

5.4.2 Themes/Evidences from Users Reaction/Perception Evaluation

From analysis of interviews conducted, seven (7) themes and twenty-five (25) categories emerged (see table 5.27 below). Below, each of the themes is discussed and linked to relevant research proposition(s) and context, as appropriate.

■ **Context/Domain and Users:**

At least, five interviewed lecturers voiced their views, either explicitly or implicitly, on the ILABS/ILAT implementation context. For instance, lecturers L1, L2, L4, L5 and L7, acknowledged the implementation of the ILABS and its product (i.e. ILAT) within the accounting domain. This is implied in their responses at different stages of the interviews. An example is the response given by lecturer L5, in which he states:

> *In the first instance, ((a bit of silence)) basically, the person must know accounting because we are talking of accounting software, then the person at a stage will largely be a student or lecturer, then the basic computer knowledge is very-very essential for it to be used, with that, then one can develop on it as it goes on.*

This confirms the supposed domain/context of the research under discussion. Equally, lecturer L5 identified the category of users appropriate for the ILABS and ILATs generated from it, as reflected in his response stated above. He claimed that the ILABS will be useful for lecturers—in that it can be used to generate ILATs that aid teaching—while the ILAT generated will be useful to students. It thus, aligns with the target of this research.

From the foregoing, it could be concluded:

- That the ILABS was actually implemented in the initially planned evaluation domain/context of the research—the numerical problem-solving context of the accounting and finance discipline.

- That the ILABS will be suitable for lecturers as a tool for generating teaching aids while ILAT will be appropriate for the students as learning aids.

Table 5. 27: Themes from Lecturers Responses of the ILABS Evaluation

Themes / Corresponding Categories	N	Lecturers who hold this idea
Context/Domain & Users of Builder/Tutor		
Implementation context	5	L1; L2; L4; L5; L7
Target users	1	L5
Characteristics & implementation of Builder/Tutor		
Features & tutoring strategies	8	L1; L2; L3; L4; L5; L6; L7; L8
Implementation of features & tutoring strategies	7	L1; L2; L3; L4; L5; L6; L8
Limitations of tutoring strategies	2	L5; L6
Impact of features on tutor behaviour	4	L2; L4; L5; L8
Learning Process & Benefits		
The learning process	8	L1;L2; L3; L4; L5; L6; L7; L8
Learning benefits	7	L2; L3;L4; L5; L6; L7; L8
Learning issues & their implications		
Observed learning issues	1	L6
Perceived learning issues	1	L6
Teaching & learning medium	1	L1
System production & boundaries		
Production capabilities	8	L1; L2; L3; L4; L5; L6; L7; L8
Production restrictions	3	L1; L5; L8
Production requirements	3	L6; L7; L8
Production time	5	L2; L5; L6; L7; L8
Usability issues		
Interface design and ease of use	8	L1; L3; L5; L6; L7; L8
Functionalities	8	L1; L2; L3; L4; L5; L6; L7; L8
Learning curve	5	L1; L4; L6; L7; L8
Builder usage requirements	8	L1; L2; L3; L4; L5; L6; L7; L8
User satisfaction	8	L1; L2; L3; L4; L5; L6; L7; L8
User-support utilities	3	L1; L5; L7
Overall ratings	7	L1; L2; L3; L4; L6; L7; L8
Builder Extension—Extending & Enhancing Learning		
Mobile accessibility	1	L3
Gaming approach	1	L3

■ **Characteristics, Implementation and Impact of the Builder/Tutor:**

As shown in the table above, all the lecturers (L1 to L8) expressed views regarding features and tutoring strategies that embody the ILABS being evaluated, which can be used to configure and generate ILATs in the numerical problem-solving context of the accounting and finance disciplines, earlier confirmed in the theme described above.

Accordingly, lecturer L8 described features that constituted the ILABS through the following statement:

> *Yah, the calculator panel, navigation guide, interactive message, close button, graph button, the marker board, the print board are there; two approaches under tutoring strategy panel—model tracing...process monitoring options—are also there. Yah, those are some of the things in it.*

Similarly, lecturer L1 describes the features at two different instances thus:

> *Yah, you have quite a number of items on the screen—like calculate, navigation panel, message panel, and various other buttons in the same group. With this design in place, at least, the window allows you to select any of these items by ticking the boxes, followed by ok.*

As well, the latter stated thus, *"Also, down the window, you have tutoring strategies with four options available."* The last two instances indicated the composition of ILABS, and they both confirm the earlier quoted statement of lecturer L8.

As part of the software design (see chapter three), the above-identified features—as constituted—were initiated to represent and implement the characteristics of the metamodel that underpins the ILABS. They are, therefore, expected to be embedded in any ILAT generated, depending on the configuration of the said tutoring system. In order to confirm whether the design works as envisaged, some of the statements of the lecturers were appraised. Analyses of interviews conducted seem to confirm the design, indicated by the affirmative comments made by all the lecturers, which pointed to the embedment of the ILABS features in the ILAT generated. This is evident in some statements considered below. For example, lecturer L3 stated thus: *"Yes.....they all appeared on the tutor... You can switch the calculator on and off, the message panel display alert message........"*. Lecturer L5 responded as follows: *"Sure, it does reflect; you can see all the options selected are already included in the sample tutor generated.........*", while Lecturer L8 gave the following response:

> *Yah, they are all well-reflected there; I can see calculator panel in the new module, similar to the old one. It is also possible to switch calculator on/off*

depending on the option you prefer to use [I: does that mean they are reflected?]. Yes, I think so. Yes..yes.....because once you turn on the calculator, the eTutor starts responding to each action.......

From the above aforementioned quotes, it is thus implied that the research design, with respect to the implementation of the metamodel, worked as planned, since the features built into the ILABS were replicated in the ILAT generated from it. Although those responses confirmed the replication of the ILABS features in the ILAT, not much or detail was given about the implementation and behavioural context of the features in the ILABS and/or ILAT, which is an important measure of the true and actual implementation of the metamodel under evaluation. This aspect, which was also revealed by some other statements, is considered next.

As mentioned in above quotation, lecturer L1 stated that the features are implemented in a template form in the ILABS, so users just tick any of the options desired to embed in the ILAT generated. Also, the tutoring strategies were implemented in such a way that they provide four optional selections. The foregoing-described implementation is fully explained in another statement by the latter lecturer—lecturer L1:

....it supports two main strategies ((referring to tutoring strategies)), but it gives you four options. You can create a tutor with either of the strategies, or with both, you can also decide not to use any of the strategies. In that case, tutor generated will not provide guidance during learning. So, they are the four options available, you can use any of them to generate a tutor, I think it makes sense, you know.

In addition to the above statement, lecturer L6 described the implementation:

the::::e model tracing approach [I: Yah], and the process monitoring approach [I: yah]. Em::;m, and then one could also use both approaches, or you may decide not use any of them.

Both latter quotes described how the tutoring strategies—a class of features—were implemented within the ILABS, but did not indicate how they were implemented in the ILAT. Nevertheless, it thus means that any of the options can be selected to drive or guide the behaviour of the ILAT generated. The implication therefore, is that the intelligent tutor behaves or responds according to the strategy embedded or configured into it.

In order to investigate the implementation of the features within the ILAT generated, the remark of lecturer L8 is handy, thus:

> *Both of them ((referring to Byzantium ITT and sample tutor)) don't allow you to enter any value directly when the calculator is on; you either use the calculator buttons or pick from the boxes on the interface, except when the calculator is off.*

Similarly, lecturer L6 stressed the implementation, role and consequence of not using the calculator:

> *Well, I don't know whether it is a weakness, just what I mentioned the other time, some students prefer to use their own calculator, you know, and once they do that am:::m, it affects one of the strategies, you know, em::::m... one of the eTutor strategies only work when you use its calculator.*

The last three responses reflected both the implementation and role of the calculator in the ILAT. From the quotes, it shows the calculator determines the data entry mode, as well as the functioning of the tutoring strategies. It therefore indicates that there is a sort of exchange between the user and the system, in the form of a bi-directional communication, resulting in feedback whenever the user action is found faulty or wrong at any step or stage during learning. This can be translated or linked to one of the concepts examined within the metamodel, i.e. the conversation concept. Hence, it could be said that the generated tutor imbibes conversation, in the form of bi-directional communication, to achieve its tutoring goal, through inputs into the system and corresponding feedback from the system. It also presupposes that the tutoring strategies are linked to the calculator. It thus indicates implementation design that was adopted in this work. It also points to the important role of calculator's state, with regard to the implementation of the tutoring strategies and the consequential behaviour of the ILAT.

In summary, it can be concluded that calculator's state plays a strategic role in the implementation of the features within the intelligent tutor. As the quotes indicated, the calculator impacts the data entry mode and the tutoring strategy route that can be taken during learning via the ILAT. When in the "ON" state, it enforces the use of the calculator as medium of learning and provides the user with two optional tutoring routes—either model tracing or process monitoring route—for the ILATs that employ a dual tutoring strategy option; but when "OFF", users are allowed to use their own calculator and only the model-tracing route can be taken.

As identified by lecturer L6 in above latter statement, and also noted in another instance, the statement below —in a sense—strengthens and confirms his earlier observation of the ILAT's workings and the consequential tutoring restriction, due to non-use of the calculator, thus:

> *Really, is not......is not......is not, the:::e...the::::e only, what I can note there, is that em:::m.....am::::m, some students prefer to use their own calculator [I: Ok]. If they can't use the system calculator what that means is that em::::m the::::e process-monitoring approach cannot be followed [I: Ok]. So, I don't know whether that em:::::m......// excuse me.....so am::::m.......the:::e, I don't know to what extent that can affect the students, but that's the only am::::m adverse remark I think em:::m = [I: Ok, that it restrict students to system calculator] = it restrict students to use the system calculator. Some of them prefer to use their own calculator. I don't know why, but my experience, they would tell you they don't want to use the system calculator; they want to use their own calculator.*

Thus, it can be confirmed that the above observation was real, i.e. non-use of the calculator by a student restricts the functioning of one of the tutoring strategies. Although the interviewee may see this as a weakness or limitation of the tutoring strategy employed, it can be said that the design was deliberate. The so-called limitation can be associated with the implementation design adopted in the research, aimed at monitoring the cognitive process of students during learning. The cognitive process is a key concept being evaluated, likewise is the conversation concept mentioned above; they both constitute the metamodel under consideration, and addressed within the numerical problem-solving context of the accounting and finance disciplines. Thus, the online device, i.e. the calculator, can be said to have been designed around the tutoring strategies to achieve the aforementioned research objective. From another perspective, the observation and other evidences above can be said to be a confirmation that the design and implementation of the cognitive visibility and conversation concepts worked as planned, since they both manifested in the behaviour of the ILAT generated from the ILABS.

Furthermore, on the role or impact of the ILABS features on the ILAT generated, more evidence from the interviews shows that the embedded features did actually impact the intelligent tutor's behaviour. This is clearly stated in a statement made by lecturer L8:

> *It displays under the tutoring strategies panel, model tracing, process monitoring, both strategies, and none options....it allows you to tick only one. I can see that when you select the process monitoring approach...e:::m..there is something new, not in the old Byzantium version, which is there......tutor behaves differently from the old version. For instance in the old version, the tutor prompts whenever there is a mistake but it prompts only when you put in the final value, but from the new version, I can see that the tutor monitors my activities, it prompts at every stage. So that it doesn't wait until you get to the final value before it prompts.*

From the foregoing statement, mention was made of the tutoring strategies supported by the ILABS. When one of the strategies—the process monitoring—was selected, it impacted the behaviour of the ILAT generated thereafter. It enables step-wise monitoring of learning activities, as implied or elucidated by the above statement, indicating the role or effect of the selected strategy on the intelligent tutor. It thus implies that the process monitoring approach aids identification of cognitive nodes in a numerical problem-solving context, since each solution step or unit is monitored, resulting in an alert whenever a wrong step or misconception is identified. Thus, it confirms the alignment of the ILABS/ILAT to one of the metamodel concepts: the cognitive visibility concept. From a theoretical point of view, if the cognitive process of a student can be made visible, it is assumed that misconception can be detected and appropriate feedback can be provided to enhance learning. Hence, it can be concluded that the implementation aligns with this assumption, judging from the above statement. This conclusion can also be implied from the statement made by lecturer L5, which described the system feedback time in relation to the calculator's state, thus:

> *From my own little experience, I think it does ((context of probe question—tutor monitors solution steps)). It does because the timing of the messages depends on the status of the calculator ((whether ON/OFF)). As earlier stated,*

once the calculator is on, it enforces correction of every wrong step; otherwise you can't move forward. But, when you don't use the calculator, it will only enforce correction if the final answer is wrong. In my own layman understanding, I assume it monitors the solution steps and it guides you through to the final solution.

Conclusively, all the above suggests that the research pedagogic metamodel can be implemented in ILABS. This stance was inferred from the various evidences provided above, indicating the implementation of features that stand in for the theoretical concepts within the ILABS, and reflected through the reaction or behaviour of the ILAT generated from it. Also, the position was taken because the implementation occurred within the numerical problem-solving context of accounting and finance, as identified in the earlier discussed theme, as the context of the research discussed in this thesis. Hence, the proposition 1.1, which states that "a *metamodel can be conceptualised and implemented in an ILABS,"* was thus confirmed from the evidences provided by the interviewees' responses. It does so because ILABS supports features that enables the generation of tutoring systems underpinned by its underlying assumptions within the numerical problem-solving context of the accounting and finance discipline.

The above stated evidences also indicated that dual tutoring strategies—i.e. model tracing and process monitoring—can be implemented within an ILAT. This position was taken, drawing on the sample ILAT generated and utilised in this evaluation, and the comments made by interviewees in that regard, indicating the behaviours of the intelligent tutor with respect to the calculator's state and each of the strategies. Hence, it can be concluded that proposition 1.2, which states that "*It is possible to generate tutoring systems that support process monitoring and model-tracing from the implemented metamodel.*" was thus confirmed, moreso that the ILABS provides two tutoring strategies, which were implemented in four different ways as described in the evidences cited above, and found in the raw text extracted from the interview transcripts.

■**The Learning Process and Benefits:**

Regarding this theme, all the lecturers' views appear to align as demonstrated in the excerpts from the interviews transcripts that were analysed. These views cut across various concepts that can be regarded as elements of a learning process. Referred elements emerged from the implementation of the ILABS features in the sample ILAT generated from it. For instance, lecturers L2, L4 to L8 affirmed that the ILAT enabled bi-directional communication between the learner and the system via inputs and feedback, believed to have occurred in an interactive way. In support of the aforementioned, lecturer L5 noted as follows,

> *It provide feedback based on learning actions and status of the calculator; for example, if an illegal action is taken—maybe you picked a wrong variable or operator—when the calculator is on, the system alerts you immediately; otherwise, it allows you to carry on with your work. So, I see it as responding to user actions in an interactive manner."*

In order to reinforce one of the concepts in the foregoing statement, it should be noted that lecturers concurred that the intelligent tutor gives feedback in response to users actions, either immediately or delayed. The timing of the feedback depends on the tutoring approach route taken and the calculator's state applicable at the time of learning. These two factors impacted when feedback is given. It thus implies that the sample ILAT provides two types of feedback depending on the factors mentioned. A position also expressed by lecturers L1, L3, L6, and L8. Two other lecturers also upheld the view expressed in the above quote, regarding the sample ILAT's interactive learning capability. In order to demonstrate this, lecturer L7 voiced her view, saying:

> *Sure, it does; the learning tool provides feedback because if a student putsin a wrong figure the system tells him it is wrong and he cannot make any progress. You have to stop or try again. So, in that way, both the learning tool and the student are engaged in a form of interactive session; not just a reading session alone, which textbooks just give, because of that, most students won't know how to go back and get the problem solved.*

The latter was very emphatic of the benefits that can accrue from the interactivity of the tutoring system, to the extent of comparing it with a learning situation involving only a textbook. This position strengthens the inclusion of interaction as part of a learning process, and it thus confirms its inclusion in the metamodel being evaluated. This concept in combination with the bi-directional communication concept, constitute what is regarded as conversation, a learning medium treated in this research. It assumed no meaningful learning can occur without interaction and communication. It should, therefore, be noted that interaction may not necessarily be physical, it may exist in other ways; for instance, communication in a virtual world of computing.

Further to the above, lecturer L6 described the learning process enabled by the tutoring system, thus,

> *......, but then, it also monitors—you know—the process,....[I: OK]....you know, and that is a very good advantage. So once you are able to monitor the process, the:::e....it engages the students—you know—in their em::m learning activities by providing feedback appropriately.....[I: so, in::::::indirectly it::::it engages students in conversation in form of feedback].....feedback, exactly....[I: in conversation].....yes; then, it monitors the steps the students are taking to arrive at their answer, [I:....to arrive at their solution] yes, so it can say, ok, at this step you are wrong.....not necessary at the final answer.....*

The above view identified some further concepts that the ILAT promotes. Concepts such as monitoring, engagement, and implied conversation between system and learner were identified. Regarding the monitoring aspect of the intelligent tutor, the above quotation demonstrated that the tutoring system provides two types of monitoring activities: step-wise monitoring of learning, and goal-oriented monitoring. Step-wise monitoring implies that ILAT monitors the cognitive nodes or problem-solving steps of a learner. In contrast, goal-oriented monitoring refers to the monitoring of the end-result of a problem being solved. In essence, the ILAT can be judged to support both conversation, as well as cognitive visibility of the learning process.

As described by lecturer L6, the ILAT monitors the learning process, identifies incorrect step(s) and notifies the learner immediately. Therefore, it demonstrates the intelligent tutor was able to identify the point of misconception; it aids early remediation, since an incorrect action was detected at step level, and appropriate feedback was provided—as noted in the participant's remark. This same view was shared by all other lecturers. For example, lecturer L3 describes early detection of misconception, thus:

> *Tutor responds in different ways.......turn on the calculator,........ Whenever you try to derive a variable and you pick value or click button it does not expect, it instantly gives a message, it does not wait until you drop the final value. It monitors every step you take to derive a variable, for instance. But,*

the case is different when calculator is off, it only tells you after deriving a variable—whether its correct or wrong,...ea:::a and you can drop or enter value directly into the spreadsheet without any hindrance. So, that is how the sample tutor behaves at different situations.... I think the calculator is controlling how it works, more or less........

While lecturer L1 describes the detection of misconception and remediation as follows:

And the tutor responds according to the status of the calculator. You have the option to use the calculator or not; but if you use the calculator, that means you turn it ON, then all your activities must be through it and it monitors them. Any instance of error at any stage of your work, the system informs you immediately. So, you have to correct the step before you can be allowed to carry on. In my opinion,it seems to be observing each step as you go on.

The immediate two quotations correlate early detection of misconception to the use of the calculator—a learning medium. It thus demonstrates and strengthens the significant role that the calculator plays in achieving these two aspects of the learning process.

In addition, lecturer L6 believes that the ILABS enhances comprehension by providing a tutoring strategy option, which when selected, enables the generation of an ILAT that does not provide tutoring guidance during learning. Through such means, learners can be evaluated, i.e. comprehension of topic learnt can be examined. He thus expressed this stance by saying that this feature is:

......is a good one, because if you don't use it, it means you want to test their knowledge [I: OK], in an examination scenario; [I: scenario?] you need that [I: OK]. You don't need any of the em:::m tutoring approaches [I: OK], so that they can do it on their own without being led by the system [I: Ok].

Similarly, lecturer L5 aligned with the position, which was expressed in his idea,

As far as I am concerned, this builder has a simple design, uses simple language.....as well as other attributes I just mentioned. Tutors from it generate practice problems, which students can use to test their pulse, which is essential anyway.

Thus, from the phrase "*.....generate practice problems, which students can use to test their pulse,.....*", learners can be examined through practice questions, thereby enhancing comprehension.

As evinced above, regarding the practicability of implementing a metamodel in an ILABS, the research participants added by making known some learning benefits accruable from such implementation. At least five of the participants believed that the implementation provides some learning benefits. Lecturer L5 noted that the implementation will aid the evaluation of learning through provision of practice problems, as mentioned earlier. Lecturer L6 said the implementation, which resulted in varying ILABS features, enables the control of learning, enhances detection of guessing and aids evaluation of comprehension, via disabling/enabling tutor properties. Likewise, lecturer L7 observed that the calculator aspect of the implementation aids access to tutoring guidance, enhances computation of accounting variables during learning, and also encourages interactive learning sessions. As a result, L7 claimed that it makes learners competent in problem solving.

Added to the above-stated learning benefits, L8 noted that the process monitoring approach enhanced step-wise learning, and will enable learners *"...to think back"* (i.e. enhance reflection), as reflected in statement made: *"I don't think so; it ((i.e. Process monitoring approach)) would help them to learn every step, to think back at every step."* Consequently, it established "reflection" as a key element of a learning process. More so, it was promoted as a key learning method by the CA theory—discussed in chapter three of this thesis (Brown, Collins, & Duguid, 1989; Collins, 1991a; Collins, 1991b; Collins, Brown & Holum, 1991), which happens to be a constituent of the metamodel being evaluated through the ILABS. Accordingly, the theory uses "think back" or "reflection" as the medium to deepen learning or comprehension (Collins, Brown & Newman, 1989; Collins, 1991a; Collins, 1991b), and the metamodel being evaluated has been described in this light. It therefore validates the implementation carried out in this work. Also, L8 noted that the implementation of dual tutoring approaches provides complementary learning benefits. This, once again, confirms that model tracing and process monitoring approaches can be jointly implemented within a tutoring system. Therefore, proposition 1.2, which states, *"It is possible to generate a tutoring system that supports model-tracing as well as process monitoring from the implemented metamodel,"* is hereby confirmed further.

Hence, from all the above, it can be concluded that ILABS enables the production of ILAT that evolves a learning process, which enhances learning. This claim was established through participants' comments, from which elements of a learning process were established. These include interactive and bi-directional communication (thus, enabling conversation between learner and system), monitoring (which enables cognitive visibility through step-wise monitoring; and/or goal-oriented monitoring of end-results), misconception, feedback and remediation. It also promotes evaluation of learning to enhance comprehension. It can also be observed that interactivity, bi-directional communication and monitoring of learning activities provide a platform for other elements to evolve within the tutoring system. Thus, it further confirms workability of the fundamental assumptions of the metamodel to underpin an ILABS that can be used to generate tutoring systems in the numerical problem-solving context. It also establishes the possibility of implementing two tutoring strategies within an ILAT. Lastly, it demonstrates that conversation and cognitive visibility are twin-elements that enhance other elements of the learning process, thereby enhancing learning and the evolvement of a viable learning process.

■**Learning Issues and Implications:**

As part of the evaluation, one participant identified some issues that relate to learning via tutoring systems. These issues relate to the learning behaviour/attitude, preference and feelings of students, while learning a numerical subject. Accordingly, lecturer L6 said:

> *And most students, what they usually do, is to memorise the steps and once they go into it, em:::m they just—you know, regurgitate what they have memorised into the system.*

The quote reflects some students' behaviour towards tutoring systems. According to the interviewee, instead of using such systems for active learning, students do otherwise. Such a learning attitude was attributed to the unchanging or static nature of such tutoring systems. However, with the development of the ILABS, the interviewee believed that such behaviour will be eliminated or reduced drastically, since the ILABS enables reconfiguration of tutoring systems generated from it. It thus provides an advantage, because the tutor's interface can be modified, thereby curbing or reducing this negative learning attitude. This position was expressed in another statement L6 made:

For instance, if you.....em::::m, a student might memorise a step—you know, in one of the modules, and the system monitors the steps, if the students goes wrong, the programme tells the students.......and the students would try another thing, not necessary that the students knows it, but you know, by trial and error the student makes it. This one ((i.e. the new module)) you can disable that, so that, if the student gets it wrong once, you know the student is just guessing and that is better; em::m, it is an advantage.

Also, the latter observed that students prefer to use their calculator, either in an attempt to spend the minimum time with the tutoring system, or because they feel the tutoring-system-based calculator wastes their time, or because they are used to their own calculator. A clear reason for this attitude could not be identified. This suggests that such students were externally motivated to use the tutoring system, possibly to satisfy academic requirements. Notwithstanding, it should be noted that the ILABS embedded a calculator into ILATs as part of the process of implementing the process-monitoring approach. Where such a tutoring route is not required, the calculator can be excluded from the ILAT during configuration. Hence, the ILABS provides a lot of flexibility in addressing some of the concerns identified. All said, it could be stated that learning requires learners to be motivated, to achieve learning gains, instead of being coerced in order to fulfil academic requirements or for any other reasons. In response, the ILABS enhances this aspect of learning by enabling configuration and reconfiguration of tutoring systems, and deepens comprehension through provision for the generation of practice problems and a viable learning process, as evinced in the above quote and other views voiced by participants.

■Production and its Boundaries:

The prototype ILABS provides features that enable the production of ILATs, and their variants through reconfiguration or modification of existing ones, as opined by all the interviewees. Five participants opined that ILABS is not restricted in terms of the number of ILATs that can be produced, and all of them asserted that it is capable of producing variants of a tutoring system, when needed. For example, lecturer L1 provided a detailed description of the production capabilities of the ILABS in the following way:

With the features in place for now, you can create tutors and you can modify them as well. I don't see the builder tied to a specific topic, like the marginal *costing module we generated. Since features used for marginal costing sample module are reusable, so, I think it can produce tutors for different topics, unless a topic requires something not provided, which for now I cannot think of. You see, for now, it has several objects that can be used to create boxes for different variables in a spreadsheet-like screen. It allows you to label them also, set different attributes, such as the position of variables on the screen, specify their sizes and so on. You have basic arithmetic operations too, such as addition and subtraction, which are basically what we need in accounting. When you create a tutor, you can still modify it later. I think with all these, it's fine for now.....it is flexible enough to use.*

The view expressed, apart from addressing the production of an unrestricted number of tutoring systems, equally touches upon some other key issues of the ILABS, including its reconfiguration capability. The implementation domain/context was implicitly mentioned through the phrase, "*....which are basically what we need in accounting",* further confirming the domain/context of this research.

Accordingly, L3 also held a similar view, which was expressed thus:

....builder has enough features, to the best of my exposure, that can be used to produce eTutors for different topics of accounting....am really looking forward to having a copy installed on my computer. What attracts me most is its adaptability, you know. You can create new tutors or modify an existing one to suit your purpose. It allows you to move around the objects on a tutor interface, increase or decrease their size, change colour.....that it self-suffices for now. So, the builder affords different ways, to illustrate lessons to my students since I can generate several of it, and I can also modify any one generated. For me, the features suffice for now.....though, there can be room for improvement. At least, let's have this for now....we can then think of improving it later when need for it arises. It's really going to make things easier for lecturers, you know, because with it, we can produce tutorial aids.

The above two opinions provide detailed insight into what is achievable with the ILABS. As described, none of the interviewees had any contrary view to the production capabilities of the ILABS; instead, each reinforced its capabilities in varying ways. The only foreseeable or future restrictions relate to the limited number of operators and features that complex problems may require, which are not provided by the ILABS. These views were expressed by three lecturers, L1, L5 and L8. Consequently, the explanation that can be given relates to the need of accounting problems/topics, which this ILABS catered for. Any other requirement may then be considered in future expansion work. Outside those issues, the ILABS was very well appraised in terms of its production capability.

As a follow-up to the latter points, three of the interviewees voiced opinions on production requirements that should be met to produce an intelligent tutor. Lecturers L6 and L7 observed that whenever an intelligent tutor is modified, regeneration of such a tutor will be required to effect the changes made. This view was express by L7 as follows:

> *Sure, it does; the builder allows you to move objects; you can also remove any of the tutor elements. But, I observe you have to regenerate the eTutor after modification to effect the changes.*

Also, L8 expressed her own view on the production requirement thus:

> *As you can see, the builder tool provides some features; once they are selected from the template.....automatically, they are included in the tutor, although you have to specify how they appear and where to place each feature on the interface, like the position and size of the calculator, message box, boxes for cost price, sales price, quantity, revenue and so on.......example is this sample tutor. So, you can generate tutor very easily with it and.............*

Basically, the above quotes indicated two main requirements: regeneration of ILATs to effect changes, and specification of attributes of assets constituting a tutoring system. Furthermore, four participants were of the view that the ILABS enhanced production time; although they affirmed that the ILABS speeds up production, they predicated the time required on some factors, which include problem complexity and pre-configuration plan.

■**Usability Issues:**

As presented in table 5.27 above, authors' (i.e. lecturers') usability ideas of the ILABS touches on several aspects of the usability measures. These ideas were described in varying ways; they cut across various usability aspects such as functionalities, interface design, ease of use, learning curve, user satisfaction, user-support utilities and overall ratings of the system. Also, lecturers addressed requirements that should be met for a user to be positioned effectively to use the ILABS. Some of the comments are presented below. Apart from two instances, where two lecturers—L6 and L8—mentioned a functionality bug, all the lecturers positively appraised the usability of the ILABS. For instance, L7 described the builder as follows:

> *Sure, you can achieve your goal with this tool, it has a simple design that makes things easy, and the buttons are responding accordingly. I think it is working fine.*

In same light, L3 said the following:

> *It's not ambiguous....[I: what do you mean by ambiguous?]...very easy to understand the terms used. You see:::e and it does not require a lot effort to use....ea:::::a....it does not require technical expertise to understand, the interface is very simple and straight forward. With this, so::o I don't have to waste my time before I can use the system.*

Lecturer L1 equally added voice to the usability description of the ILABS by saying:

> *I don't see any difficulty with this tool at all. The design is simple enough for anyone to understand and use. Although, like any other new product, you may need someone to introduce its workings or get a manual to explain how to use it, on this, I will recommend that a user manual, to accompany it. If this can be done, it will be okay, because it has a simple and easy to use interface, honestly; very simple design to be precise. So far, everything about it is working fine.*

While the above quotes positively appraised the ILABS by describing the quality of the various usability aspects—the functionalities, interface design, ease of use, and so on, the latter noted the need for a user manual or human support to enhance usage of the tool. This latter point touches on the learning curve of the ILABS, on which four lecturers provided varying remarks about. Lecturer L1 believed that human support or a manual may be required at first instance of use, as indicated in the immediate above quote; L6 commented that extensive training is not required, but effective use of the builder will improve with familiarisation; while L8 described the learning curve with a conditional remark thus:

> *I don't foresee any problem at all; I mean, it doesn't take long to learn how to use it; as long as, just like I said earlier on, as long as you are computer literate, you know how to click a button and you can read and you know when to click OK and when not to, I mean, it has a simple outlook, very straightforward.*

In line with the condition stated above— the need to be computer literate, it should be noted that this position was also held by other participants. They believed computer literacy is a basic requirement to use the builder; that no extra computer skill/knowledge, such as programming skill, is required. As a matter of emphasis, L5 said: "*you don't need to be a programmer before you can use it ....yes, you don't need to be a programmer.*"

Also, all the participants concurred that, since the tool was developed for the domain of accounting, discipline knowledge will be required to generate useful tutors via the ILABS. On this, L5 noted that only basic accounting knowledge will be required; L6 said a fair knowledge of accounting is a precondition; while L3 noted that vast knowledge of accounting is a precondition to easy configuration of meaningful accounting tutors, "*....accounting knowledge must be there, he must be very vast in.....ea:::::a....accounting, in all aspect of accounting before he can easily use this......*". All-in-all, the various remarks on knowledge/skill satisfied the condition "*....and by non-programmers (lecturers)*" of proposition 1.2, which states thus: "*Implemented metamodel can be used to create unrestricted number of tutoring systems within short span of time and by non-programmers (lecturers).*" It thus concludes the fulfilment of the four aspects of the proposition touching on ILABS capability to generate: unrestricted number and variants of tutoring systems, short production time, and knowledge/skill requirement.

Generally, on the usability of the ILABS, all the participants expressed positive satisfaction with the level of the development being at its prototype stage. This was expressed in varying ways. Accordingly, lecturer L3 said:

> *I so much admire it....this programme you are developing;.......If all these innovations can be injected into the educational system......., I believe we*

shall not be in this state. So, I implore you to bring this effort....this innovation into reality....into wide use......fast track the innovation, so that we can change the system.

Additionally, the overall ratings were above average; the lowest rating was 75 percent, accorded by lecturer L7 thus: "*I will give it about 75% ((referring to usability of ILABS)).*"

With above positive and varying remarks, made with respect to various usability dimensions, it can be concluded that the ILABS was a successful implementation of the metamodel. The tool was appraised usable for the purpose and persons designed it for. Hence, the research proposition 1.4, "*users of implemented metamodel have a positive perception about its ease of use and usability*", can be said to have been satisfied.

■Builder Extension—Extending & Enhancing Learning:
Notwithstanding the above described achievements, one of the lecturers made certain recommendations towards extending the usefulness of the ILABS. Accordingly, L3 raised two vital developmental areas that should be considered, to achieve high students' patronage of tutors generated from the builder, "*Yes, now.* ↑ *Yes! Yes, apart from making it available on the Internet,......it should be extended to mobile or any other mobile device, so that students can learn through it.*" Equally, the latter said:

Mobile devices....O:::o, it will. It will enhance learning. When they do it like games on mobile devices, you know....they love games....when they play it like games. You see::e, I want to see marginal costing.....if marginal costing is this, what will be the revenue...if revenue is this, what is the cost of this.....you see, like a game.....they would....that will interest them........, you see, it will just be at their fingers tips when you ask them questions.

According to the quote, extension of the ILABS to the two identified areas—i.e. extension to mobile devices and generation of game-like tutors—would widen access to tutoring systems; as well, enhances learning. Acknowledging their viability, this research could not accommodate them within the current research scope; hence, they were identified as possible future research areas.

5.5 Discussion of Findings

The above quantitative and qualitative analyses provided insights into issues on production of tutoring systems within the numerical problem-solving context. It demonstrated how, from a theoretical ground, a practical platform can be evolved to aid learning of a numerical domain, hitherto seen as problematic.

In this section, outcome of the research work, which encompasses design, implementation and evaluation, using mixed methodology, is presented and discussed. Due to adopted evaluation methodology, synthesis of findings from the quantitative and qualitative approaches were made, and forthwith discussed in the light of previous studies in the literature.

5.5.1 Findings Related to Propositions:

In this sub-section, discussion of the findings from empirical evaluation in respect of the four propositions drawn from research question one is presented thus:

■**Proposition 1.1:** A metamodel can be conceptualised and implemented in an ILABS.

In chapter three, ACCAM—the metamodel utilised in this research—was conceptualised based on two learning theories, CT (Pask, 1976a; Scott, 2001a; Boyd, 2004; Heinze, Procter, & Scott, 2007) and CA Theory (Collins, 1991a; Collins, Brown & Holum, 1991; Dennen, 2004; Dennen & Burner, 2008). Characteristics of the ACCAM were identified and discussed in the context of the research to aid subsequent development of ILABS in line with the proposition addressed here. Also in chapter three, implementation of the ACCAM was discussed. ACCAM assumptions were translated into features provided by ILABS. These two research efforts aimed at satisfying the above stated proposition on one ground.

On the other hand, an evaluation to confirm the implementation of the pedagogic metamodel was carried out. This was undertaken to validate the alignment of the ILABS to the assumptions of underlying ACCAM. Findings show that the metamodel was successfully implemented in the ILABS. It was accomplished by translating the key theoretical assumptions—conversation and cognitive visibility—into visual assets and strategies. These were then implemented as main features of the ILABS, thence confirmed the design described in chapter four. Also, findings indicated that these features were embedded in ILATs generated. This was confirmed through the visual appearance of some of the assets. Equally, they were observed through the behavioural patterns exhibited by the intelligent tutors, which conforms to the expected behaviour of the theoretical assumptions, as discussed in previous chapters. None of the participants from both methodological approaches, held a contrary view regarding the design and the implementation; instead, participants concurred that ILABS explained the behaviour demonstrated by the ILATs generated from it. In same light, no external factor(s), demographic or computer characteristics, impacted views raised.

The qualitative aspect of the analysis added deeper insight into what was revealed. It brought to light key features of the ILABS, on which the theoretical assumptions were hinged. Thus, it demonstrated that theoretical assumptions can be translated to implementable features. Accordingly, it identifies that the calculator's states ("ON" or "OFF"), had impact on both data entry mode and the behavioural patterns of tutors during learning. None-use of the calculator, as observed by participants, disabled one of the tutoring routes during learning, thus, preventing mapping of the cognitive process of a learner, but it enabled conversation between learner and system. This observation, therefore, stimulates thoughts towards other design methodologies that could be employed, in order to implement cognitive visibility, independent of a learning medium or device, such as the calculator. However, none-use of the tutor's calculator, in this case, which disables cognitive mapping of a learning process, should not be seen as weakness of the implementation adopted, rather, a deliberate design that was implemented, to take advantage of the main tool (i.e. calculator) usually used—in practice—in the domain of current research.

The above implementation of the ILABS assumptions, using a combination of process-monitoring algorithm and ILAT's calculator, demonstrated the viability to reveal cognitive nodes, although different from other approaches in the literature (see VanLehn, 1988; Pena & Sossa, 2004; Zarandi, Khademian & Bidgoli, 2012). For example, Zarandi, Khademian & Bidgoli (ibid) utilised a fuzzy-expert system to implement cognitive mapping in an ITS. Although the approach was not implemented in the context/domain of current research, results showed that the experiment provided individualised instructions based on a learner's educational status in consonance with the purpose of the research. Similarly, Blessing et al. (2009) authoring work also confirms that cognitive nodes can be mapped. This, therefore, indicates that cognitive nodes, when accurately mapped, could be used to guide learning; this aligns with the assumption that underlies ACCAM—the current research's metamodel. However, unlike other approaches mentioned, the approach adopted in current research has the potential to capture all problem-solving steps, once taken via the calculator. It shows its viability to reveal cognitive nodes while solving a numerical problem. These are interpreted by an appropriate algorithm, and then used to provide appropriate feedback based on the learner's current learning situation, similarly confirmed in Zarandi, Khademian & Minaei-Bidgoli (2012). Another advantage of current implementation is that it provides learners an extra learning route, not provided by above stated approaches, whenever they choose to opt out of the cognitive mapping route.

Having said that, the above-mentioned research findings provide insight into how ACCAM translates into practice. It indicates that practise, especially when it relates to education, should not be separated or distanced from theory; theory should drive it and determine its characteristics and outcome. This insight tends to address Self's (1990b) argument, in which the latter claimed that many research works in the field of ITS/Authoring lack formal theory, or at best, only make claim to informal theory. For example, Gilbert et al. (2011) adapted an open-source engine for an intelligent tutoring system to provide training within a game-engine-based synthetic environment. Despite the achievement claimed, there was no link to any theory, either formal or informal. So, features constituting such an authoring tool cannot be linked to educational theory; this limits/restricts its educational values, when considered from a theoretical ground. In contrast to the latter, Blessing et al. (2009) developed a tool that enables authors who are not cognitive scientists or programmers to create a cognitive model in a context/domain outside the current research's context/domain. This work was linked to its root, Anderson's ACT theory of cognition. This work, although drawing from cognitive science, strengthens the position taken in this research that theory should shape the development of an educational tool. Hence, it can be argued that an educational tool should have theory-bearing, to determine its learning objectives, educational expectations and learning boundaries. Also, the theory background should shape and determine features that should be incorporated in such an educational tool in order to accomplish the theoretical assumptions that underpin it.

Considering the above, this research aligns with the idea of Self (1990b) on the need to formalise the development of educational tools through formal theoretical underpinning. It does so by treading the path of theory-to-practice, which some current works attest to or, at least, identifies with theory (Blessing et al., 2009; Zarandi, Khademian & Minaei-Bidgoli, 2012). Despite their alignment with Self's (1990b) idea, it should be noted that they were implemented in the domain/context outside that of current research. Neither did they emerge from combined learning theories, as done in this research. Hence, current research thus overcomes some theory-practice issues such as theoretical assumptions and their manifestations, unlike previous studies in the ITS/Authoring field without any theoretical background. It also contributes to research in the field of ITS/Authoring research, specifically in the numerical problem-solving context of applied numerical domains (e.g. accounting), which for now, lacks comprehensive theory-based ITS authoring research, to the best knowledge of the researcher.

■**Proposition 1.2**: It is possible to generate tutoring systems that support process monitoring as well as model-tracing from the implemented metamodel.

Through the quantitative and qualitative evaluation methodology, certain understanding was gained with respect to ILABS' capability, in terms of its implementation of dual tutoring strategies within a tutor and its attendant implications and benefits. Both empirical studies revealed that ILABS support features that can generate ILATs with multiple tutoring strategies. In achieving that, it provides four optional tutoring routes: model tracing; process monitoring; dual tutoring-strategy (mode-tracing and process monitoring combined); and no strategy route. While the quantitative analysis confirmed the implementation of both tutoring strategies, the qualitative aspect of the research gave a detailed description of how to achieve it. It indicated the tutoring options that can be explored. Thus, the qualitative corroborated the findings from the quantitative, but provided in-depth understanding of the implementation and implications. Accordingly, it was realised that the implementation of either the dual tutoring-strategy or the process-monitoring approach, requires embedment of the calculator in the tutor, being instrumental to its implementation.

Findings suggested that the dual tutoring-strategy provides flexibility during learning. It promotes both step-wise and goal-oriented monitoring of the learning process. Implementation enables switching between two strategies constituting the dual-strategy, therefore enabling users to opt for any of its strategies with its attendant learning implications. While the process monitoring enhances early detection of misconception, encouraged reflection and enabled timely remediation of a misconceived step—as expressed by participants, in contrast, the model tracing component of the dual strategy detect misconception late. Consequently, the latter tends to result in misconceptions overlap. Users might not know the step, within the problem solving space where misconception occurred; as such, they might be compelled to start afresh a problem-solving scenario, in contrast to process monitoring. Accordingly, evaluators positively appraised the process-monitoring route and its option within the dual tutoring-strategy route. It shows that this tutoring strategy does have a learning edge for promoting step-wise monitoring or revealing cognitive nodes of a learning process, which thus aid the generation of appropriate feedback to guide learning. It in turn promotes reflection that is perceived to deepen comprehension, compared to the model-tracing option, therefore enhancing learning.

Critical to this research, is that findings have confirmed that multiple tutoring strategies can be implemented and thus possess a learning edge, especially in a domain/context, which lacks the current type of research. This is similar to what Alpert, Singley & Carroll (1999, p.7) observed when they said, *"multiple personified advisors may offer an advantage over a single tutor if their multiplicity successfully mirrors and conveys the categorical distinctions in the tutored domain."* In the latter, multiple agents were utilised to drive learning and were found to have a learning advantage over single-agent tutors. Similarly, it can be argued that the implementation of multiple strategies in this research provides learning advantages. Learners scan manoeuvres between strategies as convenient and dictated by their learning needs.

■**Proposition 1.3**: Implemented metamodel can be used to produce unrestricted number of tutoring systems within a short space of time and by non-programmers.

While both empirical studies confirmed the ILABS' production capability to produce an unrestricted number of tutoring systems in the numerical problem-solving context of accounting and finance discipline, the studies equally affirmed its capability to produce variants of ILATs. However, the qualitative studies provided insight into possible restrictions that may occur in the future. It recognises that some problems may require operators outside the arithmetic operators set provided by ILABS, which of course, are the basic operators usually employed in the accounting domain. It also revealed that some complex problems may necessitate additional features, not provided by ILABS, but these features could not be conceived or identified. Such issues are not unexpected in the first place, essentially, being this being the first version. They appear consistent with other ITS-authoring evaluation studies. For example, Ainsworth & Fleming (2006) evaluated an authoring tool meant to allow educators with no programming knowledge to design learning environments. Findings from the evaluation indicated that the authoring tool exceeded its initial expectations, but that improvement to its design could further enhance its functionality. Thus, it confirms that identified restrictions in current research are not out of place in a work of this nature; it would, therefore, contribute to further future work that may be required.

Notwithstanding the foregoing restrictions, the ILABS provisions were certified sufficient to accommodate basic accounting problems. This achievement, which is of primary research concern, can be judged to have fulfilled one of this research's objectives, similar to other accomplished authoring research objectives in the literature (Ainsworth & Grimshaw, 2004; Ainsworth & Fleming, 2006; Blessing et al., 2009), but in a different context. In the same light, the qualitative aspect elucidates the production requirements, hitherto, not revealed by the quantitative. It identified, the need to regenerate the tutor after modification in order to effect the changes. Likewise, it provided insights into the production process of the ILABS, which includes specification of properties for visual assets that constitute a tutor.

As per production time, the studies show that the builder enhanced production time. However, there were varying views on required production time. Findings from the quantitative analysis appear to suggest further work was necessary towards improving the turnaround time of the ILABS. On the other hand, the qualitative aspect hinged/predicated production time or turnaround time on problem complexity and pre-configuration plan. Participants were of the view that problem complexity goes a long way to determine production time. Added to that, they upheld the opinion that production requires planning ahead: planning the interface outlook, deciding the visual assets to be incorporated in a tutor, and other necessary components of a tutor. These views seem to align with recorded authoring experiences. For example, as noted in Ainsworth & Fleming (2006), authors took between six and eleven hours to author a four-hour course on "Understanding Shapes"—i.e. about three hours per one hour of instruction; whereas, when trainee teachers were presented with previously authored course to personalise for their students—it took them 90 minutes to customise a four-hour course. This difference in production time seems to portray a trend associated with authoring work, which cannot be ruled out. Hence, it suggests that complexity, volume or nature of a problem do impact production time. Notwithstanding this trend, as part of future considerations, ILABS features that might adversely impact production time should be identified/isolated and worked upon in order to enhance authors' productivity.

Analyses of responses equally indicated the knowledge/skill required to use the ILABS. Both studies revealed and confirmed that programming skill is not required. While this finding seem to align with authoring work for non-programmers, as demonstrated in Ainsworth & Fleming (2006), which indicates that it is achievable; on the other hand, Blessing et al.'s (2009) work shows that half of the people who created better cognitive models with the authoring tool could lay claim to be programmers. The implication, then, is that programming skill may impact the use of the authoring tool. That is not to say ITS authoring tool for non-programmers is not accomplishable. It only suggests that such a tool must be designed to meet the level of the target category of users, to which Blessing et al. (ibid) also agreed. In contrast, findings revealed the need for domain knowledge—i.e. knowledge of accounting. It predicated on the point that the ILABS was domiciled in an accounting domain; hence, to generate meaningful accounting tutoring systems, knowledge of the domain will be required. Qualitative responses also indicated that if the ILABS is extended to alternative numerical domains, such alternative domain knowledge will equally be required. The qualitative aspect further provided depth, by making known the computer knowledge required. It thus revealed the sufficiency of basic computer literacy in order to use the ILABS, while it upheld the need for ILABS' application domain knowledge.

■**Proposition 1.4:** Users of implemented metamodel have a positive perception about its ease of use and usability.

The ILABS, a practical manifestation of ACCAM, was positively appraised. Responses from both evaluation methodological axes attested to its ease of use and usability. These positive views cut across various segments of the ILABS, touching on the simple interface design, the ease of use, terms employed, functionalities, the template and menu-driven design adopted. However, there was variance in opinions across institutions. Other demographic characteristics and computer experience did not impact users' views. Despite the variance, the views were still positive regarding ILABS' usability.

From the quantitative analysis, one of the usability scale items indicated the need for introductory support, in order to enhance proficiency while using the ILABS. The same view was corroborated by qualitative responses; in this aspect, participants specifically identified that some type of initial support was required, as either a user manual or human support. They noted that a user manual should accompany the final product. Despite that, the need for extensive training was ruled out outright; instead, it was believed that familiarisation would improve ILABS usage. Considering the foregoing, it seemed to suggest the need to incorporate help facilities as part of the ILABS' features, in order to enhance usability. Qualitative responses equally identified occasional system freezes. This was actually due to a corrupted configuration data file, not a functional problem. Despite the explanation, it seemed to trigger thoughts into ways such issues can be prevented. Thus further work to prevent recurrence of the issue should include automating the detection and repair of corrupted data files; display of a message, such as "file unreadable"; inclusion of a recovery mechanism—to free the system from the effect of such files.

Despite the foregoing, participants strongly held the view that the ILABS was a successful implementation of ACCAM, having achieved its purpose (i.e. enabled non-programmers to generate tutors). Also, extensive training that has been confirmed, not required, seems to inform the learning curve of ILABS. It demonstrates the high value attached to the simplistic design approach adopted, and how it has impacted positively on the usability of the ILABS, by eliminating the need for extensive training, which most products often require. Also, the provision of help facilities seemed to add value to usability; it informs a usability feature that should be part of a software product, such as the ILABS, being evaluated and discussed here.

While the research acknowledges issues are identified above, previous studies are not exempted (e.g. Moundridou & Virvou, 2002a; Ainsworth & Fleming, 2006); the stated studies too had their successes, as well as issues requiring attention. For instance, Ainsworth & Fleming (2006) evaluated REDEEM—an authoring tool for declarative aspects of knowledge, which was meant for non-programmers. The latter study associated its success to the extent the authoring tool was usable by its intended users (non-programmers). Despite claimed success, it acknowledged improvements to its design were required, which could further enhance its functionality. Similarly, Moundridou & Virvou (2002a) evaluated WEAR—an ITS authoring tool for algebra-related domains. They acknowledged the need to enrich the role of authors/instructors by providing relevant information during the ITS development cycle using the WEAR authoring tool, similar to the help requirement identified in current research. Therefore, issues identified in current research do not in any way invalidate the success achieved. Instead, they constitute part of the ILABS developmental process, and usually such issues come up in evaluation studies of this nature. Furthermore, they indicate areas that can be worked upon to improve the standard of the ILABS, and to enhance its usability.

Moreover, while REDEEM was successfully implemented for the declarative aspect of knowledge, the current research was a success story in the procedural aspect of knowledge, and in numerical disciplines of accounting and finance.

Considering the above as a whole, the system seemed to conform to usability standard identified in previous authoring works and usability literature (Lindgaard, 1994; Murray, Blessing & Ainsworth, 2003; Blessing et al., 2009; Gilbert et al., 2011). As Lindgaard noted, usability can be measured by the extent a product satisfies and overcomes usability dimensions and typical defects respectively. Thus, effectiveness, flexibility, “learnability” and attitude were identified as necessary dimensions that should be satisfied; while typical defects related to navigation, screen design and layout, terminology, feedback, consistency, modality, redundancies, user control, and match with user tasks. Accordingly, the ILABS and related remarks, when considered from the stated usability window, can be argued to have satisfied most of the stated criteria, with the exception of two identified issues (the help facility and occasional freezes due to a corrupted data file). Participants’ remarks were positive on the builder’s effectiveness and flexibility; they commented positively on its unrestricted production capability, speedy production time, ease of use and its functionalities, which, according to them, enables users to achieve their authoring goal. On “learnability”, they ruled out the need for extensive training, but identified with its gradual learning curve, and the need to familiarise with the system (to improve usage). Generally, users’ attitudes were positive towards the ILABS; they felt satisfied with the tool and the level of development. They also commented favourably on its simple interface design, which touches on navigation, screen design and layout, and terminology. Overall, as an indication of their satisfaction, the ILABS was rated above average.

5.5.2 Other Findings:

Apart from findings related to the propositions addressed above, some additional insights emerged from the qualitative analysis; these are discussed below.

■Learning Process and Elements:

The qualitative responses unveil a learning path and the elements instrumental to its emergence. It reveals a pattern, demonstrated by the relationship between emerged learning elements. This relationship indicates how each element enhances the occurrence of the other. The emerged pattern, known as the learning process in this work, can be described by the following elements: bi-directional communication and interactivity, monitoring (either at step or goal level), misconception, feedback, reflection, remediation and evaluation. They were instrumental in achieving effectiveness of the ACCAM learning process and the behaviour of the ILATs generated from the ILABS, as indicated by responses from participants. The emerged learning pattern shows that learning takes place via communication between the learner and the system in an interactive manner. This process, which is constantly monitored, identifies students' misconception and/or provides feedback, as appropriate. Where misconception occurs, it enables the learner to reflect on and remedy this. Thus, it describes the process of learning via the learning tool. The learning tool further provides features that enable practice questions to be generated, encouraging evaluation of students' comprehension of domain concepts in a numerical problem-solving context. This was enabled by the provision of the ILABS.

The learning platform created further demonstrated the successful implementation of ACCAM in the ILABS, and its eventual transferability into tutoring systems. It shows how a learning process can emerge from a pedagogic metamodel underpinned by pedagogy theories, what its constituents could be, and what relationships should exist between them. It indicates that conversation (signposted by two elements, bi-directional communication and interactive learning) and cognitive visibility (represented by step-wise monitoring) are twin elements that enhance other learning elements, thereby enhancing learning and the evolvement of a viable learning process. It further proves that the implemented metamodel enables the production of ILATs that can evolve a learning process, which enhances learning.

Buckler (1996) presented a model for a business organisation. The latter acknowledged that the model's components improved learning within a business organisation, similar to the current research, which addressed learning within an educational environment. However, its success was predicated on the quality of leadership provided by managers and team leaders. On that note, it can be induced that the success of any system would be dependent on the extent of integration of its components and drivers. With respect to current research, the learning process promoted by ACCAM was effective in enhancing learning. Its success could be attributed to the twin elements of conversation and cognitive visibility, which were demonstrated to have enhanced the emergence of other elements of the learning process.

Also, the elements mentioned seem to be some of the outcomes of the learning effectiveness of the approach implemented in this research. Collectively, they demonstrated certain learning characteristics, which could be related to the attributes of meaningful learning as identified in other studies, such as active, constructive, intentional (or reflective), authentic (or contextualised), and cooperative (or collaborative/conversational) learning (Jonassen and others, 2003 as cited in Pongsuwan et al., 2011). These attributes were demonstrated by the learning process that evolved from this research as follows: it encouraged active learning because learners are involved in knowledge construction; aided constructive learning since learners partake in problem solving; enabled reflective learning since it encouraged reflection via misconceptions and remediation; learning was contextualised because it took place in a problem-solving context; and was cooperative since it promotes learning via conversation, which thereafter enhanced the cognitive visibility of learners. Thus, the learning process elements impacted each other, which consequentially contributed to effective learning that was achieved via ILAT.

Along those lines, motivation was also identified as a necessary element that can promote active learning. This research suggests that when students are motivated, the right learning attitude can be exhibited and learning can be purposeful and fruitful. The possible students' behaviour towards some features of a tutoring system, as highlighted by participants, further strengthened the importance of motivation, for learning tools to achieve their educational purpose. This requires addressing the two classic aspects of motivation: intrinsic and extrinsic (Ryan & Deci, 2000). While the fulfilment of academic requirements may contribute to extrinsic motivation, it informs the need to research features that can stimulate extrinsic motivation, which should be incorporated into a learning tool. On the other hand, the intrinsic aspect of motivation requires a change in the learning behaviour of students. All said, the ILABS features that were embedded in ILATs generated enhanced motivation, as observed by participants, but more work may be required in this area. However, this could not be considered in this work, since it falls outside its scope, but can be considered in future research due to its importance.

■**Learning behaviour:**

As shown in table 5.27 above, one of the participants observed that students demonstrate certain behaviours, such as memorising solution steps and regurgitating them during tutorial or evaluation sessions. Such learning attitudes, according to the participant, are encouraged by a static tutoring system's interface, i.e. tutors developed for a specific purpose, which are fixed—not modifiable. The implementation of the ILABS, which enables reconfiguration, disabling/enabling, and personalisation of tutoring system features, was seen as a positive development that would eliminate/curb or reduce such negative learning activities. This further confirms the learning advantage due to the implementation of ACCAM in ILABS. It also provides insights into features that can contribute to active and enhanced learning. Similarly, it highlights features that should be considered in practice, when developing a software product for educational purposes.

■Builder Extension

The qualitative aspect of this research also provided insight into other researchable areas that can enhance the usefulness of the ILABS and its products. These include mobile-based learning, and a game-based learning environment. Researches into mobile-education indicated that enhancement can be brought into teaching and learning via mobile devices (Lehner & Nosekabel, 2002; Sharples, 2002; Sharples, Taylor & Vavoula, 2010). While such devices support/facilitate teaching and learning, no attempt should be made to replace traditional education involving teachers. Instead, such devices should widen accessibility to teaching and learning platforms, supporting a traditional medium of education, as projected in this research. This position aligns with the argument in Lehner & Nosekabel (2002).

Similarly, game-based learning tends to stimulate interest in learning through fun or what can be called "serious entertainment"; serious, in the sense that it is purposeful, goal-oriented and has an educational undertone; not just for the fun of it. When conceived from this background, an extension of the ILABS towards building game-like tutoring systems can be said to be sensible. More so, that no such ITS authoring works that can generate game-like ITSs in the context and domain of this research exist, to the best knowledge of the researcher. Some of the few works in that direction—Authoring and/or Intelligent Tutoring System (Costu, Aydin & Filiz, 2009; Johnson, 2010; Li, Zhuying & Bing, 2010; Gilbert et al., 2011), although, not in the numerical problem-solving context of accounting domain, opined that game-based learning has a motivational effect, with potential to stimulate learning and enhance learning outcomes.

This thus confirms the viability of extending the ILABS into this new learning environment.

While these research areas sound viable, with possibilities of widening access to tutoring systems and motivating learning, they fall outside the scope of the current research. However, realising the role mobile devices play in the life of people in the world today (see Lehner & Nosekabel, 2002; Sharples, Taylor & Vavoula, 2010—on mobile-based learning), the interest students developed for games, and the potential of such a learning environment (see also Costu, Aydin & Filiz, 2009; Johnson, 2010; Amioa, Gardent, & Perez-Beltrachini, 2011), accounted for some researches in mobile-learning and educational game software. Thus, this strengthens their viability as possible future research areas. In order to achieve these goals, such research efforts come with their attendant implications. They require incorporating features that can enhance the production of tutors that run on mobile devices. Also, consideration must be given to software development tools that support the development of mobile applications. On the other hand, the game-like aspect will require redesigning the architecture of ILABS, to accommodate features of the gaming environment.

With the above envisaged future work, this research hoped to widen access to tutoring systems in the context/domain of this research, which for now, lacks rich and innovative tutoring systems. Equally, consideration of game-like tutors is likely to capture the interest of students, who like games, eventually enhancing their learning. Lastly, mobile and game-like environments would open up a research window to test-run ACCAM in a new learning environment not considered at the onset of this research.

Table 5.28: Summary of findings with respect to research question one:

Propositions	**Quantitative Findings**	**Qualitative Findings**	**Conclusion**
1.1 Conception and implementation of the pedagogic metamodel (ACCAM)	■Authors confirmed the implementation of conversation and cognitive visibility assumptions in both ILABS and ILAT. ■However, authors felt slightly stronger in favour of conversation assumption (bdasum1, M=4.65) than cognitive visibility (bdasum2, M=4.50). ■No independent factor influenced authors' reactions to the implementation except authoring experience which was moderately in favour of those without experience. Thus, authors with experience were more stringent in their views, although they still strongly confirmed the implementation. ■Also, authors confirmed that the implementation of the assumptions in ILABS informed their implementation in ILAT constructed (via the ILABS) since there is a strong positive correlation between the respective scales (rho=0.542, n=82, p<0.01).	■ Authors qualitative views described the features and tutoring strategies embodied in ILABS which were replicated in the ILATs constructed an indication that ILABS characteristics informed its product (ILAT). ■Authors also described the tutoring behaviour of the ILATs constructed via ILABS. The description thus indicates the presence of the theoretical assumptions of the pedagogic metamodel in ILAT, implying that ILABS implemented these assumptions which in turn produced what is observed in the ILATs.	■In chapter 3 of this thesis, the conception and implementation of ACCAM was discussed. Therein, ACCAM characteristics and how they translated into features in ILABS were elucidated. ■Authors quantitative and qualitative evaluations confirmed the implementation of ACCAM assumptions in ILABS, which in turn manifested in the behaviour of ILATs constructed therefrom.
1.2 Implementation of dual strategies—model-	■Authors strongly affirmed the implementation of dual-tutoring strategies in ILAT constructed. This suggested that ILABS implemented relevant tutoring strategies since	■Authors qualitative views indicated that ILABS made available three different tutoring routes and allowed the implementation of dual strategies in any	■Both quantitative and qualitative reaction of authors indicated the implementation of dual strategies in ILABS (an implementation of

tracing and process monitoring.	ILAT is a product of ILABS and can only be constructed based on the features provided by the latter tool.	ILAT constructed. ■Also, their qualitative views showed that constructed ILATs demonstrated different behaviour at each instance of the implemented tutoring strategies. Equally, they mentioned that authors were able to switch from one tutoring route to another within an ILAT, thus indicating the implementation of dual strategies in ILAT.	ACCAM), which in turn manifested in its product (ILATs).
1.3 Unrestricted Number of tutoring system	■Authors affirmed ILABS can be used to construct unrestricted number of ILATs covering different subject matter topics and their variants.	■Authors qualitative responses confirmed the production of varying number of ILATs. ■Qualitative views noted that ILABs is limited in terms of operators (only supported basic arithmetic operators). Thus, some complex problems may necessitate additional operators not currently supported.	■Authors noted the construction of unrestricted number of ILATs, variants. However, ILABS is noted to have limited operators which may affect construction of ILATs for complex problems.
1.3 Production Time	■Authors affirmed that ILATs can be relatively constructed in a short span of time, but they express scepticism towards production within five hours. ■The result suggested refinement/improvement to ILABS production turnaround time; however, the research acknowledged that ILABS is still a prototype.	■Qualitative views hinged production turnaround time on problem complexity and pre-configuration plan. Thus, this suggests that complexity, volume and/or nature of a problem impacts on production time.	■Qualitative and quantitative views indicated the contruction of ILATs within a short span of time but hinged the turnaround time of ILABS on problem complexity, volume of work required and/or nature of the problem to be captured.

1.3 Required Special skills	■Authors strongly express the need for prior knowledge of target subject matter to capture its ontology in ILABS. ■However, authors outrightly rejected the need for programming skills to use ILABS, suggesting that the latter tool was user-friendly to accommodate non-programmers.	■Qualitative views confirmed the need for domain knowledge in order to be able to capture the subject matter ontology in ILABS. ■Also, authors affirmed that programming skills are not required to use ILABS but noted that computer literacy would be necessary.	■ Both quantitative and qualitative views confirmed the need for domain-specific knowledge to use ILABS effectively. However, programming skills are not required, while basic computer literacy is considered sufficient.
1.4 Usability evaluation	■Authors generally agreed on the usability of ILABS having satisfied necessary ease of use and usability criteria, except that they expressed the need for introductory explanation or help facilities at first time of use. ■None of the independent factors influenced the views expressed except the nature of the institution of authors. This requires further investigation in the future.	■Qualitative reponses confirmed the ease of use and usability of ILABS but also expressed the need for help facilities to aid the use of the ITS authoring tool. Authors specifically mentioned the support required as either a user manual or human support. ■ Authors noted that ILABS does not require extensive training to use but familiarisation with tool could enhance usability. ■Qualitative reaction also noted the need to include recovery features whenever errors or problems such as file opening problems occur.	■Authors confirmed the ease of use and usability of ILABS. As a result, extensive training is not required but familiarisation is considered to enhance usability of the authoring tool. ■They also noted the need to improve ILABS to accommodate recovery routines that can prevent occasional freezes due to system or file errors.
Other findings (not addressing any of the above propositions)		■Qualitative views indicated that the twin implementation of conversation and cognitive visibility in ILABS make known other theoretical elements (e.g. bi-communication, interactivity, monitoring, feedback, misconception, reflection,	■The twin implementation of conversation and cognitive visibility in ILABS enabled the emergence of a learning pattern and identification of certain theoretical elements that can

		remediation and evaluation) that comes into play during a learning process which can contribute to effective/meaningful learning. ■Authors identified motivation as an essential attribute that can promote active learning. ■ILABS encourages learner's self-evaluation due to its features that enable the construction of ILATs that generate practice questions. ■Since ILABS enables reconfiguration, disabling/enabling of features and the personalisation of ILATs, certain negative learning behaviour (e.g. memorising/regurgitating) of learners can be eliminated or curbed.	promote effective/meaningful learning. ■Motivation is considered a key learning attribute for active learning to be achieved. ■ILABS encourage learner's self-evaluation due to support for practice problem generation features. ■ILABS enables the elimination/curbing of certain negative learning behaviour of learners.

5.6 Summary

This chapter analysed and discussed the quantitative and qualitative data with respect to the implementation of ACCAM in ILABS and ILATs generated from it. This was undertaken from the perspective of authors (i.e. lecturers, the target users). Summary of the findings with respect to the foregoing and other related propositions is provided in table 5.28 above. Thus, the findings confirmed the successful implementation of the metamodel in ILABS, which in turn impacts the sample ILAT that was generated. The following chapter, chapter six, analyses and discusses the data with respect to the product of ILABS—the tutoring system generated—from the perspective of student users.

Chapter 6: Data Analysis II

This chapter aims to provide answers to issues posed in research question two. The research question was investigated via four propositions (see chapter 1—section 1.4). The propositions relate to the research's theoretical concepts embedded in ILAT(s) generated from the ILABS (the implemented metamodel) that was examined in the previous chapter. Proposition one examined the embedment of the key theoretical concepts (conversation and cognitive visibility) in ILATs generated from ILABS. Proposition two looked at the impact of the theoretical concepts on some learning objectives (such as timing and relevance of feedback). Proposition three investigated the link between two theoretical concepts (cognitive visibility and misconception). Proposition four examined the perception of students regarding the learning effectiveness of the theoretical concepts assumed to be present in the generated ILAT.

Thus, the chapter aimed to achieve three things: one, it sought to find out learners' belief or position on the possible learning impact of the above-stated theoretical concepts; two, determine the learners' reaction towards ILAT(s) generated (from the ILABS) in terms of the embedded theoretical concepts; and three, it aimed to gain users' perception of the learning effectiveness of the concepts as embedded in the ILAT(s). In order to achieve the above stated aims, evaluation of an ILAT was undertaken within the numerical problem-solving context of the accounting domain. Note that the term "eTutor" or "intelligent tutor" refers to the ILAT generated from ILABS. Data collected via questionnaires administered to students in higher education-providing institutions, being the target users of the ILAT(s), were analysed. Student users were also observed in the course of exposure to ILAT. Findings were discussed in the light of the theoretical framework underlying this research, the research context and relevant works in the literature. Furthermore, the process taken to arrive at the findings is discussed below.

6.1 Data Collection / Preliminary Analysis

The research instrument—mainly a questionnaire with open and closed questions—was administered in three phases. Two phases were pilot studies, while the third phase was the main study. The pilot studies were undertaken purposely to test the research instrument in order to evaluate its reliability and validity. However, the analysis undertaken in this chapter only centred on the main study data; the pilot study analysis can be found in appendix 6.1.

The target subjects for this aspect of the research were mainly students from higher education institutions. Table 6.1 below shows the responses with respect to the main study data collection phase, indicating the institutions that participated. The code-naming of institutions, as reflected in the table below, was done to reinforce confidentiality.

Table 6. 1: Number of Participants in the Main Study

Study Type / Institutions		**Questionnaire**
Main study	Uni.B	151
	Uni.C	149
Total		**300**

The empirical data collected for this aspect of the research was cleaned, as recommended in Pallant (2010), by checking for data entry errors and outliers. The process was implemented by generating frequency statistics of all variables, which showed the frequency of valid cases, missing values, and minimum and maximum values. These statistics were used to determine whether the values fell within an expected range. Missing values were cross-checked with original documents to ascertain whether they were genuinely missing. The report did not show any error, neither were there outliers, and missing values were confirmed real. Thus, data entered into IBM SPSS Statistics version 19 was assumed to be the true version of the content of the administered questionnaires. This was followed by preliminary analysis aligning with the literature (Pallant, 2010). This was undertaken to explore the data and guide the analytical procedure employed. See appendix 6.1 for details on the aforementioned. Thereafter, data from the closed-ended questions of the research instrument was subjected to various statistical analyses, such as descriptive and bivariate statistical analyses. Findings from the analysis are reported below. On the other hand, responses to the open-ended questions were collated and analysed. Findings from these responses and some observations made during exposure/evaluation were also presented in this chapter.

6.2 Quantitative Analysis

This section presents the quantitative analysis of students' responses to the four propositions meant to address research question two: "*Can the learner's cognitive process be made visible to aid the generation of relevant and timely diagnostic feedback in order to enhance learning effectiveness in the numerical problem solving context?*" It aimed to gain insights into students' reaction or opinion on the research's theoretical concepts, as implemented in the tutoring system(s) generated from the ILABS. The evaluation was undertaken within the numerical problem solving context of accounting and finance modules. Below, the analyses are presented according to each proposition.

6.2.1 Proposition 2.1

The learner's cognitive process can be made visible to the tutoring system

Two questionnaire scales—Cognitive Process Visibility scale (CPVSB) and Conversation aids Cognitive Visibility (CCVSB)—were examined within this proposition. It aimed to capture student users' perception/reaction to two theoretical constructs: cognitive visibility and conversation concepts, of ACCAM. The first scale examined the feasibility of the implementation of cognitive process visibility, while the second examined the instrumentation of conversations as a medium to facilitate the first construct (i.e. cognitive visibility). Details of the two scales can be found in the eTutor questionnaire instrument (see appendix 4.3). However, analysis of the main study data for the two scales is presented below.

■General Users' Perception/Reaction:

The results of the one-sample statistics and related one-sample t-test (that represents the comparison of the research's benchmark value with the mean scores of scales and their items) are presented in tables 6.2 and 6.3 below. From the one-sample statistics generated, the overall reaction mean scores for both scales were above the current research's benchmark (3.0), where: cognitive process visibility scale (CPVSB) - (M=3.73 approx., SD=0.645 approx.), and conversation aids cognitive visibility scale (CCVSB) - (M=3.90 approx.; SD=0.718 approx.).

Also, all the scale items for both scales had mean scores above the stated benchmark. For scale CPVSB, item cpvsb7—*"responses from the eTutor were relevant to my problem solving steps"*—had the highest mean score (M=3.81; SD=0.879) while item cpvsb5—"*eTutor responses shows it accurately identified my thinking process*"—had the lowest mean score (M=3.66, SD=0.980). From the latter item, an insight into the actual ability of ILAT was provided. It reveals the imperfection of a technology-based tutoring system, demonstrated by its inability to capture the totality of the thinking process of its user. Even though it may perform at optimal level, its ability to monitor close to 100 percent the problem-solving processes of its users within a numerical problem solving context did not seem guaranteed. This was voiced in the reaction of users to item five of the CPVSB scale, as reflected in the distribution of mean scores of the scale. On the other hand, the mean scores for the CCVSB scale items were very close, and above the benchmark as mentioned earlier.

Table 6. 2: One-Sample Statistics for CPVSB and CCVSB scales

	N	Mean	Std. Deviation	Std. Error Mean
Cognitive Process Visibility scale	**300**	**3.72714**	**.645073**	**.037243**
cpvsb1	300	3.74	.935	.054
cpvsb2	300	3.75	.892	.051
cpvsb3	300	3.72	.947	.055
cpvsb4	300	3.69	.951	.055
cpvsb5	300	3.66	.980	.057
cpvsb6	300	3.70	1.024	.059
cpvsb7	300	3.83	.879	.051
Conversation Aid Cognitive Visibility scale	**300**	**3.89500**	**.717536**	**.041427**
ccvsb1	300	3.89	.802	.046
ccvsb2	300	3.90	.889	.051

Despite the above, which seems to suggest agreement on both constructs measured (i.e. cognitive visibility and conversation concepts), the strength of users' position was further investigated, in order to reach a verifiable conclusion. In that respect, results of the one-sample test, as shown in the table 6.3 below, indicated significant differences between the benchmark value (3.0) and both scales, thus:

- CPVSB: t (299)=19.524, p=0.0, and the magnitude of the difference (mean difference=0.727 at 95% CI: 0.654 to 0.800) was large (eta squared=0.56);
- CCVSB: t (299)=21.604, p=0.0, and the magnitude of the difference (mean difference=0.895 at 95% CI:0.813 to 0.977) was large (eta squared=0.61).

Also, the scales items were significantly different from the benchmark; item "cpvsb5" "*eTutor responses shows it accurately identified my thinking process*", had the lowest mean score, and was used as the basis for conclusion reached:

- CPVSB5: t (299)=11.662, p=0.0 (at alpha level 0.05), and the magnitude of the difference (mean difference=0.660 at 95% CI: 0.55 to 0.77) was large (eta squared=0.31). The foregoing indicates that the eTutor (i.e. ILAT) enhanced cognitive visibility.

The above results describe the general perception of users. Whether those perceptions are influenced by some factors, are unknown; hence, the following investigation.

Table 6. 3: One-Sample t- Test for CPVSB and CCVSB scales

	Test Value = 3.0 (benchmark)					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Cognitive Process Visibility sub-scale (CPVSB)	**19.524**	**299**	**.000**	**.727143**	**.65385**	**.80044**
cpvsb1	13.774	299	.000	.743	.64	.85
cpvsb2	14.633	299	.000	.753	.65	.85
cpvsb3	13.232	299	.000	.723	.62	.83
cpvsb4	12.503	299	.000	.687	.58	.79
cpvsb5	11.662	299	.000	.660	.55	.77
cpvsb6	11.786	299	.000	.697	.58	.81
cpvsb7	16.290	299	.000	.827	.73	.93
Conversation Aids Cognitive Visibility sub-scale (CCVSB)	**21.604**	**299**	**.000**	**.895000**	**.81347**	**.97653**
ccvsb1	19.158	299	.000	.887	.80	.98
ccvsb2	17.604	299	.000	.903	.80	1.00

■Effect of Demographic/Computer Experience Factors on Users' Perception:

Descriptive statistics of users across demographic and computer experience factors shows that users were reasonably evenly distributed over more variables than others (see appendix 6.2a). Some of these evenly-distributed variables were considered in this section to examine their impact on users' responses. This investigation was employed to arrive at a logical conclusion on the constructs under discussion.

[a] Effect of Institution/Department Factors

This research investigated the main effect of two independent variables, institution and department, on users' perception of the dependent variables, CPVSB and CCVSB scales. Their interaction effect could not be ascertained, because Lavene's test of equality of error variances was significant for both scales at alpha value of 0.05, where: CPVSB (F=2.991, Sig. 0.031); and CCVSB (F=5.356, Sig.=0.001). Since significance level was reached, it implies that the error variance of the dependent variable was not equal across groups, a requirement that should be satisfied before a two-way analysis of variance could be utilised (Pallant, 2010). Hence, the interaction effect was considered inappropriate in this respect. Due to that, only the main effects of the categorical variables on the dependent variables were examined. So, an independent samples t-test was adopted. Full results of the t-test generated from IBM SPSS statistics version 19 can be found in appendices 6.4 (for institution) and 6.5 (for department).

♦ **Effects of Institution/Department on CPVSB scale—**For cognitive process visibility scale (CPVSB), Lavene's test of equality of variance across institutional groups was not significant (F=1.571, Sig.=0.211 at 5% alpha level), but was significant across departmental groups (F=5.341, Sig.=0.022 at 5% alpha level). Due to the outcomes, t-test result for *"equality of variance assumed"* was considered for institutional effect analysis, while the result for *"equality of variance not assumed"* was utilised for departmental effect analysis. Consequently, the result for institutional effect shows significant difference between the mean scores of institutional groups, Uni.B group (M=3.954, SD=0.621) and Uni.C group (M=3.498, SD=0.587), where: t (298)=6.535, p=0.0, and the magnitude of the difference (mean difference=0.456 at 95% CI:0.319 to 0.593) was moderate (eta squared=0.13). Also, the result of departmental effect shows significant difference between mean scores reaction of accounting students and other students, thus: t (298)=2.075, p=0.039, and the magnitude of the difference (mean difference=0.152 at 95% CI: .008 to 0.296) was small (eta squared=0.01). From this analysis, it shows that students from institution Uni.B had a stronger positive view about the CPVSB scale than students from institution Uni.C. So, the institution that a student belongs to actually impacted their reaction. One explanation, though, that may require further investigation in future, is that institution Uni.B is a technical-oriented institution (so students have a practical orientation to learning), while institution Uni.C is a traditional institution. This may account for the difference in their views. On the other hand, departmental groups express almost the same level of reaction since the effect of the difference was small. Hence, it can be inferred that the course a student majors in does not count much in his/her reaction to the construct examined.

♦**Effects of Institution/Department on CCVSB scale—**With respect to conversation aids cognitive visibility scale (CCVSB), Lavene's test of equality of variance across institutional groups was statistically significant (F=8.369, Sig.=0.004 at 5% alpha level); it was also statistically significant across departmental groups (F=8.263, Sig.=0.004 at 5% alpha level). It thus means that the variance across groups is not equal. Therefore, the t-test for *"equality of variance not assumed"* was considered for both institutional and departmental effects analyses. Based on that, there was a statistically significant difference between the mean scores of the institutional groups, Uni.B group (M=3.977; SD=0.617) and Uni.C group (M=3.812; SD=0.800), thus: t (278.231)=1.995, p=0.047, and magnitude of the difference (mean difference=0.165 at 95% CI: 0.002 to 0.327) was small (eta squared=0.01). Similarly, a statistically significant difference was found between mean scores of departmental groups, accounting group (M=4.014, SD=0.654) and other depts. Group (M=3.721, SD=0.772), where: t (231.207)=3.430, p=0.001, and magnitude of the difference (mean difference=0.292733 at 95% CI: 0.124598 to 0.460869) was small (eta squared=0.05). The implication of the aforementioned is that the actual impact of the institution and department on the dependent variable was so minimal or small. Therefore, it can be concluded, with respect to the CCVSB scale, that neither the institution nor department, where students belong, influenced their reaction to the construct. Instead, users' perception should be linked to their personal acceptance or agreement to the construct as implemented in ILAT.

[b] Effects of Other Factors

The possible effects of gender, qualification and previous eTutoring experience were considered. This was undertaken to enhance the current investigation by revealing possible factors that might impact—if any—the positions taken by users on the constructs (CPVSB & CCVSB) examined within proposition 2.1.

♦Effect of Gender—In the light of the above, the gender test conducted revealed no significant difference in Lavene's test of equality of variance for the cognitive process visibility (CPVSB) and for conversation aids cognitive visibility (CCVSB) as follows: CPVSB (F=2.117, Sig.=0.147) and CCVSB (F=1.498, Sig.=0.222). Thus, it implies that the variance across groups was equal. Consequently, the t-test results for equality of variance assumed was utilised for the analysis of both scales. It indicated no statistically significant difference between mean scores of males and that of females, for each of the scales, where:

- CPVSB—male (M=3.699, SD=0.663) and female (M=3.760, SD=0.625); t (297)= -0.815, p=0.416; and the magnitude of the differences (mean difference= -0.061, 95% CI: -0.209 to 0.087) was very small (eta squared=0.002); and
- CCVSB—male (M=3.914, SD=0.689) and female (M=3.870, SD=0.755); t (297)=0.529, p=0.597; and the magnitude of the differences (mean difference=0.044, 95% CI: -0.120 to 0.209) was very small (eta squared=0.0009).

♦**Effect of Qualification—**Similarly, the test conducted to determine the effect of users educational level, in terms of their highest qualification prior to the survey, indicated no significant difference in Lavene's test of equality of variance for the cognitive process visibility (CPVSB: F=0.25, Sig.=0.875); it was significant for conversation aids cognitive visibility sub-scale (CCVSB: F=4.189, Sig.=0.042). With these results, it shows that variance across groups was equal with respect to CPVSB; but unequal with respect to CCVSB scale. Thus, results of t-test equality of means, which correspond to Lavene's test of equality of variance assumed was utilised for the former scale (CPVSB), while the one that corresponded to Lavene's test of equality of variance not assumed was considered for the latter (CCVSB). Both results showed significant statistical differences between the mean scores of the two qualification levels as follows:

- CPVSB—users with O Level qualifications (M=3.511, SD=0.634) and those with higher qualifications—A Level and above - (M=3.839, SD=0.625); t (297)= -4.278, p=0.0, and the magnitude of the differences (mean difference= -0.328, 95% CI: -0.480 to -0.177) was moderate (eta squared=0.06 approx) using Cohen's (1992) guidelines; and
- CCVSB—users with O Level qualification (M=3.743, SD=0.777) and those with higher qualifications, A Level and above (M=3.977, SD=0.673); t (177.969)= -2.583, p=0.011; and the magnitude of the differences (mean difference= -0.235, 95% CI: -0.414 to -0.055) was small (eta squared=0.04)—from Cohen's (1992) guidelines.

♦**Effect of eTutoring Experience—**On the last factor, previous eTutoring experience, Lavene's test of equality of variance revealed no significant difference for both cognitive process visibility (CPVSB: F=1.010, Sig.=0.316) and conversation aids cognitive visibility sub-scale (CCVSB: F=1.323, Sig.=0.251). This implies that the variance across groups was equal for both sub-scales. Therefore, the results of t-test equality of means corresponding to Lavene's test of equality of variance assumed was utilised for both scales in the following analyses. The independent t-test results reflected statistically significant difference between the mean scores of those with (YES) and without (NO) previous experience, where:

- CPVSB: with previous experience i.e. "YES" (M=3.937, SD=0.616) and without previous experience i.e. "NO" (M=3.526, SD=0.609); t (298)= 5.814, p=0.0, and the magnitude of the differences (mean difference= 0.411, 95% CI: 0.272 to 0.550) was moderate (eta squared=0.10 approx) in line with Cohen's (1992) guidelines; and
- CCVSB—with previous experience i.e. "YES" (M=4.051, SD=0.684) and without previous experience i.e. "NO" (M=3.745, SD=0.719); t (298)=3.772, p=0.0; and the magnitude of the differences (mean difference=0.306, 95% CI: 0.146 to 0.466) was small (eta squared=0.05 approx) using Cohen's (1992) guidelines.

6.2.2 Proposition 2.2

Cognitive visibility can be used to aid the generation of relevant and timely diagnostic feedback.

Similarly, this proposition addresses two constructs, time and relevance of a diagnostics feedback, in relation to cognitive visibility (a tutoring strategy). It aims to establish the link between cognitive visibility and the constructs (timely & relevant feedback). The metamodel under consideration, from which the constructs and tutoring strategy emerged, assumes learning can be effective if feedback is given at the appropriate time and is relevant to the learning context of the student concerned. Based on that, it was also assumed that if the cognitive process of a learner can be tracked, as accurately as possible, it can enhance the identification of required feedback, and the time it is required. Tracking cognitive process is conceived feasible in the current research context, since numerical problem-solving involves a collection of solution units or steps. So, if those steps can be vividly mapped, then they can be interpreted to aid the generation of relevant feedback. In line with the assumptions, the proposition seeks to know the perception/reaction of the student users to the generated tutoring system (ILAT). The investigation also attempted to find out the effects of demographic and computer experience on users' responses. Details of the investigation are presented in the following sub-sections.

■ General Users' Perception/Reaction

As mentioned earlier in the preliminary analysis, both scales are not normally distributed. This prompted an inspection of the means and medians of the distributions, in relation to the benchmark used for assessing the direction of users' views. Descriptive statistics show that means for both scales were higher than their corresponding median, where: timely feedback, TIMFDBK (M=3.628, Median=3.500, SD=0.691); and relevant feedback, RELFDBK (M=3.874, Median=3.833, SD=0.552)—see table 6.4 below and additional information in appendix 6.2b. Despite that, both mean and median scores were greater than the benchmark (3.0), suggesting users' agreement over both constructs under consideration—at scale level analysis, although this needs further confirmation. At item level analysis, the mean scores were above the benchmark (see table 6.4 below), also suggesting that users agree on the constructs measured.

Table 6. 4: One-Sample Statistics for TIMFDBK and RELFDBK scales

	N	Mean	Std. Deviation	Std. Error Mean
Timely Feedback scale	**300**	**3.62833**	**.691102**	**.039901**
timfdbk1	300	3.89	.847	.049
timfdbk2rvs	300	3.36	.963	.056
Relevant Feedback scale	**299**	**3.87402**	**.552195**	**.031934**
relfdbk1	300	4.11	.705	.041
relfdbk2	300	3.79	.848	.049
relfdbk3	299	4.06	.764	.044
relfdbk5	299	3.60	.894	.052
relfdbk6	299	3.61	1.022	.059
relfdbk7	299	4.08	.828	.048

The suggested users' agreement was further investigated. One-sample t-test was conducted using the benchmark value (3.0) as a test value (or basis for comparison). The result indicated a statistically-significant difference between the benchmark and the mean scores of scales and their items (see table 6.5 below). The results and strength/extent of the differences, indicated by the computed eta squared for both scales, are presented here:

- Timely feedback (TIMFDBK)—t (299)=15.747, p=0.0; the magnitude of the difference (mean difference=0.628, 95% CI: 0.550 to 0.707) was very large (eta squared=0.45); and

- Relevant feedback (RELFDBK)—t (298)=27.369, p=0.0; the magnitude of the difference (mean difference=0.874, 95% CI:0.811 to 0.937) was very large (eta squared=0.72).

From the above stated analysis, it can be concluded that users strongly agree to both constructs, meaning that the implemented cognitive visibility strategy aided the generation of timely and relevant feedback.

Table 6. 5: One-Sample Test for TIMFDBK and RELFDBK scales

	Test Value = 3.0 (benchmark value)					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Timely Feedback scale	**15.747**	**299**	**.000**	**.628333**	**.54981**	**.70686**
timfdbk1	18.265	299	.000	.893	.80	.99
timfdbk2rvs	6.477	299	.000	.360	.25	.47
Relevant Feedback scale	**27.369**	**298**	**.000**	**.874025**	**.81118**	**.93687**
relfdbk1	27.191	299	.000	1.107	1.03	1.19
relfdbk2	16.197	299	.000	.793	.70	.89
relfdbk3	24.082	298	.000	1.064	.98	1.15
relfdbk5	11.518	298	.000	.595	.49	.70
relfdbk6	10.239	298	.000	.605	.49	.72
relfdbk7	22.573	298	.000	1.080	.99	1.17

■ Effect of Demographic/Computer Experience Factors on Users' Perception

Following above conclusion, the effects of other interplaying factors were examined, to probe further reasons behind users' position(s). Factors hereby considered include institution, department, gender and previous eTutoring experience of users.

[a] Effect of Institution, Department and Gender on Users' Perception

This research attempted to investigate both the main and interaction effects of the above stated independent variables (i.e. institution, department and gender) via two-way analysis of variance. This was possible because:

i. Lavene's test of equality of error variances indicated no significant difference for both scales (TIMFDBK: F=0.784, Sig.=0.602; RELFDBK: F=1.193, Sig.=0.307);

ii. the required sample size per group, to attain 0.80 power at 0.05 alpha level, was reached, except for the group—"Uni.B-Other Depts-Male"—with 18 samples, which was three samples less than the recommended 21 per group suggested by Cohen (1992, p.158). This was considered insignificant, on the assumption that it will not bias the results; if at all, it will only affect the interaction effect of institution-department-gender, other interactions, and main effects will not be affected since they had a sample size above 21 (see appendices 6.6 and 6.7).

♦ **Effect on Timely Feedback Scale**—For timely feedback scale (appendix 6.6), results of the two-way analysis of variance to explore the impact of institution, department and gender on the scale, revealed no significant difference between groups at 0.05 alpha level, where:

- There was no interaction effect between institution, department and gender groups, $F(1, 291)=0.994$, $p=0.320$—thus, reflected by the effect size, which was very small (partial eta squared=0.003) .
- There was no interaction effect between institution and department groups, $F(1, 291)=0.856$, Sig.=0.356—thus, confirmed by the effect size, which was small (partial eta squared=0.003).
- There was no interaction effect between institution and gender groups, $F(1, 291)=0.066$, Sig.=0.797 (partial eta squared=0—confirming no effect at all).
- There was no interaction effect between department and gender groups, $F(1, 291)=1.289$, Sig.=0.257 (partial eta squared=0.004—indicating very small effect, if any).
- The main effects of each of the three independent variables—institution, department and gender—did not reach statistical significance, where: institution—$F(1, 291)=0.057$, Sig.=0.811 (partial eta squared=0—indicating no effect at all); department—$F(1, 291)=1.043$, Sig.=0.308 (partial eta squared=0.004—very small, no impact); and gender—$F(1, 291)=2.036$, Sig.=0.155 (partial eta squared=0.007—equally very small).

♦ **Effect on Relevant Feedback Scale**—Similarly, the results of the two-way analysis of variance (see appendix 6.7), to explore the impact of institution, department and gender on relevant feedback scale did not reveal statistical significance—at 0.05 alpha level—for both interaction and main effects, except for the main effect of department groups, stated as follows:

- No interaction effect between institution, department and gender groups, F (1, 290)=1.401, Sig.=0.237 (partial eta squared=0.005—very insignificant/small effect).
- There was no interaction effect between institution and department groups, F (1, 290)=3.260, Sig.=0.072 (partial eta squared=0.011—small effect).
- There was no interaction effect between institution and gender groups, F (1, 290)=0.364, Sig.=0.547 (partial eta squared=0.001—very small effect).
- Interaction effect between department and gender groups was not significant, F (1, 290)=1.352, Sig.=0.246 (partial eta squared=0.005—very small effect).
- The main effects of individual factor or variable on the dependent variable, RELFDBK, were statistically insignificant, where: institution—F (1, 290)=3.803, Sig.=0.052 (partial eta squared=0.013—small effect); gender—F (1, 290)=0.063, Sig.=0.802 (partial eta squared=0—indicating no effect); and
- there was statistical significance in the main effect of department groups, F (1, 290)=6.627, Sig.=0.011—however, the effect size was small (partial eta squared=0.022).

In line with the above analysis, it can be stated that users' positions on the two constructs examined were not influenced by any of the three factors—although department groups reached statistical significance, the difference was actually small.

■Effects of Other Factors

The possible main effects of qualification and previous eTutoring experience on the dependent variables—timely feedback and relevant feedback scales, were also considered. Independent-samples t-test was employed.

♦Effect of Qualification—With respect to the impact of qualification on both dependent variables, the outcome of the tests (see appendix 6.8) reveals that:

- Lavene's test of equality of variance indicated no statistical significance for both dependent variables, TIMFDBK (F=1.649, Sig.=0.200) and RELFDBK (F=0.458, Sig.=0.499); hence, satisfying one of the assumptions of the t-test.
- The t-test equality of means (for equal variance assumed) in relation to timely feedback, TIMFDBK, did not attain significance, t (297)= -1.061, p=0.290; the magnitude of the difference (mean difference= -0.898 at 95% CI: -0.256 to

0.077) was very small (eta squared=0.004), confirming further the insignificant effect of qualification on users' reaction to the construct.

- The t-test equality of means (for equal variance assumed) with respect to relevant feedback, RELFDBK, reached significance, t (296)= -3.380, p=0.001; however, the magnitude of the difference (mean difference= -0.225, 95% CI: -0.357 to -0.094) was small (eta squared=0.04), so qualification did not have much influence on users' perception of the construct.

♦**Effect of eTutoring Experience**—Similarly, results of the t-test with respect to the impact of previous eTutoring experience on the dependent variables (see appendix 6.9), revealed that:

- Lavene's test of equality of variance did not reach significance with respect to timely feedback scale (F=4.071, Sig.=0.045); on the other hand, it attained significance in relation to relevant feedback variable (F=1.659, Sig.=0.199). Consequently, the t-test equality of means values corresponding to equal variance not assumed was applied to timely feedback scale, while that corresponding to equal variance assumed was utilised for relevant feedback scale (Pallant, 2007).
- Result of t-test equality of means (equal variance not assumed) for timely feedback variable, did not reach statistical significance, t (289.475)=0.939, p=0.349; the magnitude of the difference (mean difference=0.075, 95% CI: -0.082 to 0.233) was very small (eta squared=0.003).
- While the result of the t-test equality of means (equal variance assumed) for relevant feedback, attained statistical significance, t (297)=4.050, Sig.=0; despite that, the magnitude of the difference (mean difference=0.252, 95% CI: 0.130 to 0.375) was small (eta squared=0.052).

So, it can be said that previous experience did not have much influence on the position taken by users with respect to relevant feedback construct, as well as timely feedback construct.

Conclusively, it can be stated that users agreed to the constructs measured in the proposition under examination. Equally, demographic and computer experience factors did not count as such, or impact users perceptions of the constructs evaluated.

6.2.3 Proposition 2.3

Cognitive visibility exposes/tracks learner's misconceptions

One of the fundamental issues, is that the metamodel addressed concerns learning misconception. During learning, learners' misconception should be visible enough for the tutoring system to proffer solution(s) or guide accordingly. Cognitive visibility, as a teaching-learning strategy deployed in this research, attempts to infer learners' misconception through monitoring of units/steps in the learning process in the numerical problem-solving context. As a result, this proposition measures users' reaction to the implementation of the strategy within ILAT(s) generated from ILABS. It also attempts to measure the ILAT's ability to identify learners' misconceptions. Both research issues were investigated using the misconception scale (MISCP) developed in this work. The following subsections provide the analyses of users' responses to the construct—i.e. misconception—in relation to the teaching strategy employed. Also, the research probed into the possible effect(s) of independent factors on users' reactions, in order to reach an empirically-substantiated conclusion.

■General Users' Perception/Reaction

As mentioned earlier, the ability of the teaching strategy to track learner's misconceptions was measured by the misconception scale (MISCP) of the eTutor questionnaire. Due to skewness of the scale, as reported in the normality assessment section (see appendix 6.1―section A6.1.4), mean and median scores were compared to determine the direction of users' views in relation to the benchmark (3.0). Additionally, mean score was utilised to determine the strength of the target users' views, as well as in subsequent analyses of the significance of such views.

An inspection of the descriptive statistics (see appendix 6.2b), shows that the mean score was greater than the median score for misconception construct (MISCP: M=3.680, Median=3.667, SD=0.67). Both scores were higher than the benchmark. At a scale level analysis, the overall mean score for the scale suggests that users' opinions tend towards the high end of the scale (i.e. 5.0). This indicates an agreement to the misconception construct being measured, although further probing would be required, to check the significance of its departure from the benchmark. Item-wise analysis equally indicated mean scores greater than the benchmark (see table 6.6—one-sample statistics).

Table 6. 6: One-Sample Statistics for MISCP scale

	N	Mean	Std. Deviation	Std. Error Mean
Misconception (MISCP) scale	**300**	**3.68000**	**.675944**	**.039026**
miscp1	300	3.72	.896	.052
miscp2	300	3.52	.883	.051
miscp3	300	3.80	.896	.052

From table 6.7 below, a one-sample t-test was conducted to compare the benchmark and mean scores of the misconception scale (MISCP), revealing a statistically significant difference, where: t (299)=17.424, p=0.0; the magnitude of the difference (mean difference=0.680, 95% CI: 0.603 to 0.757) was very large (eta squared=0.50). Similarly, mean scores of the scale items reached statistical significance when compared with the test value (the benchmark). With the large magnitude difference/effect, it can be concluded that users strongly support the claim made in the proposition under consideration.

Table 6. 7: One-Sample Test for MISCP scale

	Test Value = 3					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Misconception scale	**17.424**	**299**	**.000**	**.680000**	**.60320**	**.75680**
miscp1	13.983	299	.000	.723	.62	.83
miscp2	10.139	299	.000	.517	.42	.62
miscp3	15.466	299	.000	.800	.70	.90

■Effect of Demographic/Computer Experience Factors on Users' Perception

As a follow up to the above investigation, the impact of demographic and computer experience variables on users' perception was examined. This probing includes the following independent variables: institution, department, gender, highest qualification and previous eTutoring experience.

[a] Effect of Institution, Department and Gender on Users' Perception

Two-way analysis of variance was employed to probe the main effect and the interaction effect of three independent variables on the misconception scale. Lavene's test of equality of error variances indicated was not significant (F=1.376, Sig.=0.215). As a result, one of the assumptions required to use two-way analysis of variance was satisfied. Also, the required sample size per group to attain 0.80 power at 0.05 alpha level was reached, except for one group ("Uni.B-Other Depts-Male") that had 18 samples, instead of the recommended 21 per group (Cohen, 1992, p.158; see appendix 6.10).

♦**Interaction Effect**—Results from two-way analysis of variance at 0.05 alpha level revealed that:

- There was no statistical significance in the interaction effect of institution, department and gender on the misconception scale, F (1, 291)=1.081, p=0.299, and the effect size was very small (partial eta squared=0.004).
- Statistical significance was not attained for the interaction effect of institution and department on misconception, F (1, 291)=0.938, p=0.334, and the effect size was equally very small (partial eta squared=0.003).
- The interaction effect of institution and gender on misconception did not reach statistical significance, F (1, 291)=0.548, p=0.460, and the effect size was also very small (partial eta squared=0.002).
- With respect to department and gender, statistical significance interaction effect was not attained, F (1, 291)=1.578, p=0.210, and the effect size was very small (partial eta squared=0.005).

♦**Main Effect**—Also at 0.05 alpha level, two-way analysis of variance did not indicate a statistically significant difference for main effects of the variables on the misconception scale, except for the institution variable, where:

- There was no statistically significant difference for the main effect of department, F (1, 291)=0.010, p=0.920, and there was no effect at all (partial eta squared=0.0).
- There was no statistical significance for main effect of gender, F (1, 291)=0.0, p=0.998, and there was equally no effect (partial eta squared=0.0);
- However, there was statistical significance for the main effect of institution, F (1, 291)=14.122, p=0.0; however, the effect size was small (partial eta squared=0.046). Hence, the effect cannot be said to have any meaningful impact on users' views.

So, with the above results, one can conclude that institution, department and gender **variables did not reasonably impact on users views with respect to the misconception** construct measured.

[b] Effects of Other Factors

The effects of qualification and previous eTutoring experience on the dependent variable, misconception scale (MISCP), were considered.

♦**Qualification**—The results of independent-samples t-test with respect to qualification variable revealed as follows:

- Lavene's test of equality of variance indicated no statistical significance for the misconception scale, MISCP (F=0.840, Sig.=0.360); therefore, one of the assumptions of t-test was satisfied confirming its applicability in this case.
- The t-test equality of means, corresponding to equal variance assumed, indicate that there was significant difference in scores of students with O Level qualification (M=3.524, SD=0.701) and students with higher entry qualification(s), A Level and above (M=3.759, SD=0.652); t (297)= -2.867, p=0.004; however, the magnitude of the difference (mean difference= -0.235 at 95% CI: -0.395 to -0.074) was small (eta squared=0.03). This suggests that, higher qualifications did not really have meaningful impact on users' views.

♦**eTutoring Experience—**Similarly, the results of the t-test with respect to the impact of previous eTutoring experience on the dependent variables shows that:

- Lavene's test of equality of variance did not reach significance with respect to the misconception scale (F=0.389, Sig.=0.533). Thus, the t-test equality of means values corresponding to equal variance assumed was subsequently utilised for the t-test analysis.
- Result of t-test equality of means (equal variance assumed) indicated statistical significance for students with previous experience, "YES" option (M=3.803, SD=0.682) and students without experience, "NO" option (M=3.562, SD=0.651); t (298)=3.127, p=0.002; despite that, the magnitude of the difference (mean difference=0.241, 95% CI: 0.089 to 0.392) was small (eta squared=0.03).

Consequently, it can be concluded that views with respect to the construct under consideration, were not influenced by any of the factors treated. Hence, users' agreement to the construct can be said to be an assessment of what the ILAT offers them. As a result, the proposition under consideration was accepted.

6.2.4 Proposition 2.4

Cognitive visibility enhances learning effectiveness

As reviewed in chapter 2, ITS research attempts to improve learning. Deployment of cognitive visibility, as a tutoring strategy in this research context, aims to aid the achievement of the stated learning objective. This proposition therefore assumes that if cognitive visibility is implemented in a numerical problem solving context, effective learning can be realised. The analyses in this section of the thesis aim to determine users' perception of the ILAT in relation to the assumption. Also, investigation in this section sought to know the underlying belief of users, regarding cognitive visibility in relation to learning effectiveness objective. Both analytical objectives were measured through the learning effectiveness (LNEFTV) and cognitive visibility and learning (CVSBLN) scales respectively. The effects of some independent variables on users' responses to the scales, both main and interaction effect, were equally examined.

■General Users' Perception/Reaction

In line with the above, mean and median scores of respective scales were compared due to their skewness. Table 6.8 below revealed that the mean score (3.906) for learning effectiveness (LNEFTV) scale was less than its median score (3.917); while the mean score (4.048) for cognitive visibility and learning (CVSBLN) scale was greater than its median score (4.000). Nevertheless, mean and median scores for each scale were higher than the benchmark value (3.0), which is utilised to determine the direction of users' views. All the scores suggested that users views tend towards the high end of their respective scale, thus indicating agreement on the construct(s) measured by each scale.

Table 6. 8: Statistics for LNEFTV and CVSBLN scales

		Learning Effectiveness scale	Cognitive Visibility & Learning scale
N	Valid	292	296
	Missing	8	4
Mean		3.90582	4.04797
Median		3.91667	4.00000
Std. Deviation		.537694	.621204
Skewness		-.252	-.616
Std. Error of Skewness		.143	.142
Kurtosis		-.436	.423
Std. Error of Kurtosis		.284	.282

In order to confirm the suggested agreement, a one-sample t-test was employed: to determine whether there is significant difference between benchmark and mean scores of the scales; and to investigate the magnitude of the difference, if any. The results show there was significant difference between the mean scores of the scales and the benchmark, where:

- Comparison of the mean score of learning effectiveness scale (M=3.906, SD=0.538) with benchmark value (3.0) reached statistical significance, t (291)=28.787, $p < 0.05$; the magnitude of the difference (mean difference=0.906, 95% CI: 0.844 to 0.968) was very large (eta squared=0.74).
- Similarly, comparison of the mean score of cognitive visibility and learning scale (M=4.048, SD=0.621) with benchmark value (3.0) reached statistical significance, t (295)=29.029, $p < 0.05$; the magnitude of the difference (mean difference=1.048, 95% CI: 0.977 to 1.119) was very large (eta squared=0.74).

Table 6. 9: One-Sample Test for LNEFTV and CVSBLN scales

	Test Value = 3.0 (benchmark)					
					95% Confidence Interval of the Difference	
	T	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Learning Effectiveness scale	**28.787**	**291**	**.000**	**.905822**	**.84389**	**.96775**
lneftv1	31.371	298	.000	1.258	1.18	1.34
lneftv2	25.606	299	.000	1.143	1.06	1.23
lneftv3	17.893	299	.000	.937	.83	1.04
lneftv4	19.055	296	.000	.976	.88	1.08
lneftv5	18.561	298	.000	.896	.80	.99
lneftv6	19.812	299	.000	.997	.90	1.10
lneftv8	9.613	297	.000	.554	.44	.67
lneftv9	17.549	298	.000	.819	.73	.91
lneftv10	7.972	297	.000	.453	.34	.56
lneftv11	20.184	298	.000	.960	.87	1.05
lneftv12	18.591	298	.000	.960	.86	1.06
lneftv13	18.598	297	.000	.946	.85	1.05
Cognitive Visibility & Learning scale	**29.024**	**295**	**.000**	**1.047973**	**.97691**	**1.11903**
cvsbln1	27.199	298	.000	1.130	1.05	1.21
cvsbln2	23.767	298	.000	1.080	.99	1.17
cvsbln3	20.724	298	.000	1.020	.92	1.12
cvsbln4	20.717	298	.000	1.003	.91	1.10
cvsbln5	21.770	295	.000	.997	.91	1.09

Apart from the scales, the mean scores of both scales' items equally reached statistical significance, as presented in table 6.9 above. The above analysis, therefore, confirms the conclusion earlier reached, that the implementation of cognitive visibility enhances learning effectiveness. The strength of users' reaction was indicated by a large difference between compared values. Thus, it means users' views were very strong in favour of the assumption evaluated. Also, the belief of users regarding the likely learning effectiveness of cognitive visibility, if implemented in a tutoring system, tallied with users' reaction evaluation. That means, theoretically, cognitive visibility (as a construct) and the teaching strategy underpinned by it, align with the assumptions made in the metamodel that underlies current research.

■Effect of Demographic/Computer Experience Factors on Users' Perception

Further to the above investigation, the effects of demographic and computer experience factors on users' opinion, with respect to the two scales in question, were investigated. Factors considered included institution, department, gender, qualification and previous eTutoring experience.

[a] Effect of Institution, Department and Gender on Users' Perception

In order to probe the effect of above stated factors, two-way analysis of variance was employed. The analysis involved the examination of the main and interaction effects of the independent variables (i.e. institution, department and gender)—as applicable—on the dependent variables, learning effectiveness (LNEFTV) and cognitive visibility and learning (CVSBLN). With respect to the learning effectiveness scale and the interaction effect of the three independent variables, Lavene's test of equality of error variance indicated significant difference, (F=4.796, Sig.=0.0); implying a violation of one of the assumptions of the two-way analysis of variance. However, the interaction effect of two variables (institution and gender) on the dependent variable (LNEFTV) did not reach a significant difference (F=0.607, Sig.=0.611), thus was investigated. Equally, the required minimum sample size (21 per group) to attain 0.80 power at 0.05 alpha level was exceeded for all the groups involved in this analysis (Cohen, 1992, p.158)—see appendix 6.11 for details.

On the other hand, two-way analysis of variance was appropriate for the interaction effect of the three independent variables with respect to cognitive visibility and learning scale, since Lavene's test revealed no significant difference (F=1.479, Sig.=0.174). Also, the required minimum sample size per group (21 per group) to attain 0.80 power at 0.05 alpha level was exceeded (Cohen, 1992, p.158), except for one group—"Uni.B-

Other Depts-Male" that had 18 samples (see appendix 6.12 for details). Despite that, the analysis was undertaken since the difference was minimal.

♦**Effect on Learning Effectiveness Scale—**For learning effectiveness scale, tests of between-subjects effects at 0.05 alpha level revealed that:

- The interaction effect of institution and gender on learning effectiveness was statistically insignificant, F (1, 287)=1.441, Sig.=0.231, and the effect size was very small (partial eta squared=0.005).
- The main effect of institution on learning effectiveness did not reach statistical significance, F (1, 291)=0.034, Sig.=0.854, and the effect was not noticeable (partial eta squared=0.0); and
- the main effect of gender on learning effectiveness was not statistically significant, F (1, 291)=0.539, Sig.=0.464, and the effect size was very small (partial eta squared=0.002).

♦**Effect on Cognitive Visibility & Learning Scale**—With respect to cognitive visibility and learning, tests of between-subjects effects at 0.05 alpha level indicated that:

- the interaction effect of institution, department and gender on cognitive visibility and learning did not reach statistical significance, F (1, 287)=0.196, Sig.=0.658, and the effect size was very small (partial eta squared=0.002);
- the interaction effect of institution and department on cognitive visibility and learning, equally, did not attain statistical significance, F (1, 287)=0.061, Sig.=0.805, and there was no noticeable effect (partial eta squared=0.0).
- In the same light, statistical significance was not reached with respect to the interaction effect of institution and gender on cognitive visibility and learning, F (1, 287)=1.562, Sig.=0.212, and the effect size was very small (partial eta squared=0.005).
- Also, the interaction effect of department and gender on the dependent variable did not achieve statistical significance, F (1, 287)=0.623, Sig.=0.431; notwithstanding, the effect size was very small (partial eta squared=0.002).
- The main effect of institution on the dependent variable did reach significance, F (1, 287)=4.131, Sig.=0.043; yet, the effect size was small (partial eta squared=0.014).

- The main effect of department on the dependent variable equally reached significance, F (1, 287)=11.908, Sig.=0.001; the effect size was small (partial eta squared=0.04). Note that the effect of department on learning effectiveness was examined using the t-test in the subsequent section due to the significance of Lavene's test on the three independent variables (including department).
- However, the main effect of gender on cognitive visibility and learning, the dependent variable, did not attain significance, F (1, 287)=0.840, Sig.=0.360; its effect size was very small (partial eta squared=0.003).

Drawing from the above analysis, it was clear that, even where significant difference was attained, the effect of the independent variables on the dependent variables were small. So, it can be concluded that the independent variables, examined here, did not have much influence, if any, on users' views or reaction to the constructs measured. Therefore, it can be stated that the position voiced via responses to the questionnaire scales, emerged from their personal experience with the ILAT evaluated and their belief.

■Effects of Other Factors

The effects of qualification and previous eTutoring experience on both learning effectiveness and cognitive visibility learning scales were examined in this section. Similarly, the effect of department on learning effectiveness (LNEFTV) scale was examined in this subsection. In order to achieve the aforementioned, independent-samples t-tests were employed; outcomes of which were presented below.

♦Effect of Qualification—With respect to qualification variable, independent samples t-test revealed that:

- Lavene's test of equality of variance indicated no statistical significance for the learning effectiveness, LNEFTV (F=2.508, Sig.=0.114), and cognitive visibility and learning, CVSBLN (F=1.068, Sig.=0.302); it thus implies that variance across groups was equal, satisfying one of the assumptions of the t-test.
- The t-test equality of means, corresponding to equal variance assumed, did not indicate significant difference for both scales, where:
    - LNEFTV: t (289)=0.956, p=0.340—the magnitude of the difference (mean difference=0.064 at 95% CI: -0.068 to 0.196) was very small (eta squared=0.003);

- CVSBLN: t (293)=0.946, p=0.345—the magnitude of the difference (mean difference=0.072, 95%CI: -0.078 to 0.223) was very small (eta squared=0.003).

The results suggest that higher qualifications did not really have any meaningful impact on users' views with respect to either scale.

♦**Effect of eTutoring Experience**—Similarly, the results of the t-test with respect to the impact of previous eTutoring experience on the dependent variables revealed the following:

- Lavene's test of equality of variance did not reach significance with respect to the dependent variables: learning effectiveness, LNEFTV (F=0.043, Sig.=0.836), and cognitive visibility and learning, CVSBLN (F=0.030, Sig.=0.863); indicating that variance across groups was equal, which satisfies one of the assumptions of the t-test. Consequently, the t-test of equality of means, corresponding to equal variance assumed, was applied to both scales.
- Results of the t-test equality of means (equal variance assumed), with respect to the impact of eTutoring experience on the dependent variables, were statistically significant, where:
  - ➪ LNEFTV: students with previous eTutoring experience, "YES" option (M=4.016, SD=0.519) and students without previous eTutoring experience, "NO" option (M=3.798, SD=0.535); t (290)=3.528, p=0.0; magnitude of the difference (mean difference=0.218, 95% CI: 0.096 to 0.339) was small (eta squared=0.04).
  - ➪ CVSBLN: students with previous eTutoring experience, "YES" option (M=4.156, SD=0.572) and students without previous eTutoring experience, "NO" option (M=3.941, SD=0.651); t (294)=3.025, p=0.003; magnitude of the difference (mean difference=0.216, 95% CI: 0.075 to 0.356) was small (eta squared=0.03).

Consequently, the results suggest that users' views, with respect to the two constructs (LNEFTV & CVSBLN), were not meaningfully impacted by previous eTutoring experience despite it attaining statistical significance. Thus, users' agreement to the constructs can be seen as users' assessment of what the ILAT offers and their belief in viability of cognitive visibility to enhance learning.

♦**Effect of Department on Learning Effectiveness**—On the effect of department on learning effectiveness, results of the t-test shows that:

- Lavene's test of equality of variance indicated was statistically significant, LNEFTV (F=10.266, Sig.=0.002), implying that variance across groups was not equal, this value of t-test equality of means, corresponding to equal variance not assumed, was the option applicable in this case.
- The t-test equality of means (equal variance not assumed) did not indicate statistical significant difference, LNEFTV: t (228.762)=1.835, p=0.068; the magnitude of the difference (mean difference=0.120 at 95% CI: -0.009 to 0.249) was small (eta squared=0.01).

The above result, therefore, shows that department did not have any meaningful impact on users' views with respect to the learning effectiveness of the ILAT. Thus, users' views were the subject of their personal assessment of the tutoring system's provisions, in terms of its learning enhancement capability.

6.3 Qualitative Analysis

Below, some responses to the open-ended items of the questionnaire instrument and some observations during the evaluation session are reported.

6.3.1 Responses to Open-ended Questions

Some of the students addressed the items in the open-ended section of the questionnaire instrument. Although the percentage of responses was low, this research was obliged to report the remarks, since an interview was not conducted for this group of evaluators. Generally, the students indicated their satisfaction with ILAT. Also, there were some comments on the weakness of ILAT and areas that needed to be improved. Some of these remarks which provided insights into the perception of students with respect to the strength of ILAT are presented as follows:

- Makes the user think of the problem, and stimulates the mind.
- Responding to each action to avoid misconception in the process.
- It had an alternative twist to learning, which may be beneficial for students who like alternative ways.
- helps learners understand the problem without having to join a class.
- Easy to understand.
- User-friendly, easy to know about the operation process

Also, some remarks were made with respect to ILAT weakness, and areas needing improvement are as follows:

- Solving problem is just limited to certain level.
- Feedback should give formula immediately when wrong.
- Sometimes, it cannot answer what you want to know.
- Use more graphics
- Layout should be improved.
- Provide more colour, sound, etc
- The eTutor can give audio feedbacks which will be more interactive

Also, during the evaluation, students' reactions were noted. Many of the students were happy with the features and functionality of the ILAT. They all started well and progressed with the tutoring system. However, a few students were not comfortable with the frequency of the feedbacks given; at a point, they turned off the process-monitoring learning route. They felt the feedback hindered their thought process. These students prefer to explore with ILAT than to be guided. They switch to the model-tracing route, which only provides goal-oriented feedbacks. These actions seemed to contrast with the comment of some students that prefer immediate feedback, as highlighted above. What can be drawn from the foregoing is that at an early stage, frequency of feedback appears appropriate, but as learning advances, feedback should fade out according to the knowledge state of the learner or only be provided when requested.

Overall, the above indicated that ILAT enables knowledge construction, identifies misconceptions and provides feedback. It also shows that ILAT promotes reflection since it did not provide the answer or formula immediately. It scaffolds feedbacks based on attempts made by the learner, forcing reflection and remediation. This also aligns with the qualitative aspect of research question one, as detailed in chapter 5. The above also showed the need to improve on the multimedia aspect of the tutoring system generated and on the interface.

Table 6.10: Summary of findings with respect to research question two

Propositions	Quantitative findings	Qualitative findings	Conclusion
2.1 ***The learner's cognitive process can be made visibile to the tutoring system.***	■Learners affirmed their thought process was made visible since system feedback related to their problem-solving steps. ■They indicated that the totality of a learner's thought process can be visible if learners captures all their problem-solving steps through the ILAT-based calculator provided, otherwise it may be impossible to track even if ILAT performs at optimal level. ■Learners' learning-orientation impacted evaluation such that those from practical-oriented institution held a stronger view in		■Post-evaluation confimed the visibility of the learner's cognitive process. ■The medium of a learning process impacted the extent of the visibility of a learner's cognitive process. ■Learning-orientation proved to impact evaluation of the cognitive process.

	favour of the visibility of learners' thought process than those from traditional institutions.		
2.2 ***Cognitive visibility can be used to aid the generation of relevant and timely diagnostic feedback.***	■Learners strongly agree that cognitive visibility aided the generation of timely and relevant feedback.	■Learners' reaction indicated that ILAT responded appropriately to each problem-solving step, suggesting the provision of relevant feedback. ■Feedback should give formula immediately a misconception occurs. ■Frequent feedback appeared appropriate for early learners, but as learning advances, feedback should fade out according to the knowledge state of the learner. Also, the system should enable advance learners to query/diagnose their misconception whenever	■Cognitive visibility proved to aid the generation of timely and relevant feedback. ■Usefulness of frequent feedback is dependent on learners' knowledge state.

		needed.	
2.3 ***Cognitive visibility exposes/tracks learner's misconceptions.***	■Learners confirmed that cognitive visibility enhanced the detection of misconception.	■Learners confirmed that ILAT feedback prevented entrenching misconception, which if allowed to be carried forward could affect understanding.	■Cognitive visibility proved to enhance early detection of misconception and prevented entrenching misunderstanding.
2.4 ***Cognitive visibility enhances learning effectiveness.***	■Learners' belief and post-evaluation reactions suggested that cognitive visibility could enhance effective learning.	■Post evaluation reactions indicated that ILAT enhanced reflection, stimulated the mind and aided understanding of subject matter, thus indicating the learning effectiveness of the construct investigated.	■Cognitive visibility has the potential of enhancing effective learning.
Other findings		■Learners indicated that learning will be enhanced if the mulmedia features of ILAT can be improved.	■Mutltimedia features (audio, text, video) can enhance learning.

6.4 Summary

The above discussion provides the quantitative and qualitative view of target users with respect to the four propositions posed to answer research question two. Findings from the quantitative and qualitative investigation are summaried in table 6.10 above. From the analysis, it was evident that users affirmed the four propositions in context. The constructs under consideration—cognitive visibility, misconception, feedback and learning effectiveness—were strongly confirmed to be present in the ILAT. This, therefore, suggests that ACCAM's constructs were embedded and implemented in the ILAT evaluated. Thus, it indicates that the ILABS, which represents the practical implementation of ACCAM, achieved its purpose. The results in this chapter further strengthen the perspective of lecturers on the implementation of ACCAM. It also confirms that there can be an interplay between theory and practice, as noted by Hartley (2010); also, that educational tools can have a bearing on formal learning theories, unlike several research works in ITS/Authoring field that did not have any link with theory, or at best, had a link with informal theory as claimed by Self (1990b).

Chapter 7: Conclusions

7.1 Overview of the Research Aim/Objectives

This research set out to formalise the design of an ILABS, which is meant to construct intelligent learning activity tools for learning applied numeric disciplines (e.g. accounting, engineering, etc.), especially the procedural aspect of knowledge. Realising the practical nature of this knowledge involving manipulation of a numeric model of reality, the research was contextualised within the numerical problem solving of applied numerical domains. Also, the learner's knowledge construction process was key to learning (Quinton, 2010). Thus, a system was devised that makes the process visible through engagement of the learner in conversation within a cognitive apprenticeship framework. To actualise, the twin application of the conversation and cognitive visibility—through PM—was conceived as a possible means of achieving learning. In this research, PM augments conversations taking place in a CT and CA based framework to achieve improved cognitive visibility due to the limitations of CT and CA (as detailed in chapter 3). PM was implemented via an interface—a calculator—in order to bring the learner's thought process to the surface. A calculator was employed on realising that numerical problem-solving involves a lot of cognitive tasks. In turn, these cognitive tasks involve manipulation of data and this activity can be captured via a calculator.

Moreover, individual instances of ITS—tagged ILATs in this thesis—have been proven effective in enhancing learning, especially in the numerical disciplines (Kinshuk, Patel & Russell, 2000; Ritter et al., 2007; Arroyo, Royer & Woolf, 2011). Hence, the implementation of conversation and cognitive visibility in ILATs could provide support for learning procedural knowledge in such domains, if made available to cover a very wide range of topics covered within a subject discipline such as financial accounting, management accounting, investment evaluation, corporate finance, taxation, as well as numeric domains of other disciplines, such as various branches of engineering. The implementation was accomplished through an apriori link between pedagogy theories—that supports conversation and cognitive apprenticeship strategies—and an ITS authoring tool.

As a result of the foregoing, the research set out to achieve four objectives, as stated earlier in chapter 1 (section 1.3). The referred objectives pose two distinct research phases—conception/implementation and evaluation. First, ACCAM was conceptualised and implemented in ILABS—as detailed in chapter 3 of this thesis. Thereafter, ILABS was evaluated by authors in designing an ILAT which in turn was evaluated by learners—based on the research design detailed in chapter 4—and findings discussed in chapters 5 and 6 respectively. Section 7.2 provides insight into the conception and implementation phase, while sections 7.3 and 7.4 present a summary of the evaluation process and asummary of the key findings with respect to the evaluation phase of the research objectives.

7.2 Conception and Implementation of the Metamodel

In order to address research objective one, a metamodel—ACCAM—was conceptualised. It was based on two learning theories, CT (Pask, 1976a; Scott, 2001a) and CA (Collins, Brown & Newman, 1989; Dennen & Burner, 2008). CT and CA were chosen due to their support for knowledge construction that involves learning by doing (as extensively justified in chapter 3). Such a method of knowledge construction is in tune with the nature of this research context/domain—a numerical problem-solving context of applied numerical disciplines—that requires practising lots of problems. These theories (i.e. CT and CA) constituted the theoretical platform for the implementation of ACCAM in ILABS. The latter (i.e. ILABS) was utilised to produce tutoring systems that engage learners in conversation within a cognitive apprenticeship framework in order to make their cognitive process visible. Thus, the research established a formal (theory based) design approach through the implementation of ACCAM in ILABS—addressing the research aim, the formalisation of an ITS authoring tool, as discussed in chapter one. The conception/implementation of the pedagogic metamodel is therefore considered a novel design approach, in that no such ACCAM-based ITS authoring work has been undertaken. Neither has CT and CA jointly nor explicitly underpinned the design of any ITS authoring tool in the past. Moreover, none has been undertaken specifically in the numerical problem-solving context of applied numerical domains.

As mentioned above, each of the theories contributed to ACCAM. From CT, ACCAM benefited from the following key concepts:

- ⇨ Learning medium—that learning takes place through the medium of conversations about subject matter, which occurs in an attempt to make knowledge explicit. Conversation can be verbal or non-verbal (Scott, 2001b; Scott & Cong, 2010). But the non-verbal information exchange was considered in this research—as discussed in chapter 3, while the other aspect (i.e. verbal conversation) is considered for future work.
- ⇨ KR scheme—the subject matter of any learning process can be represented as entailment structures or mesh, which exist in a variety of different levels depending on the extent of relationships displayed. This was considered as a network of interrelated concepts (nodes or variables) of a target domain, as demonstrated in chapter 3.

The above contributed to the metamodel in terms of inclusion of the teaching and learning interaction (i.e. conversation), which enables knowledge exchange between the learner (the student) and system (the domain expert). The KR scheme provided insight into how domain knowledge can be captured in the metamodel and implemented in ILABS. As the theory preaches an entailment structure, it was captured as a set of rules, built from data (i.e. learning) objects consisting of variables, digits, operators, and system-defined functions. Also, the augmentation of conversation and implementation through an interface enabled the externalisation of learners' understanding of the domain in context. This, in a sense, demonstrates "teachback" of the declarative knowledge of the target domain during problem-solving. Teachback is a learning strategy embraced in CT which indicates that learning is taking place (Scott & Cong, 2010).

Although, the above CT concepts plays a key role in the metamodel conception/implementation, the cognitive process of learning, in which a novice (i.e. learner) learns from his/her master in a situated context as postulated in CA (Collins, Brown & Newman, 1989), were equally significant and prime to this research. The theory builds on traditional skill learning, in which the skill is open. In the light of the foregoing, CA proposes some teaching and learning methods, such as modelling, scaffolding, fading, etc., that can enhance learning from activity to abstraction. Also, making visible the cognitive skill that is hidden (or internal) in both the teacher (master) and the learner (student) was considered essential to achieve learning. However, Collins, Brown & Newman (1989) noted that teachers do not have access to the cognitive problem solving taking place in learners, thus making it difficult or impossible to adjust - their application of skill and knowledge to problems and tasks. Likewise, learners do not have access to the processes of carrying out cognitive problem solving in the teacher.

Thus, improving the visibility of the cognitive process of a learner was considered a key element of the metamodel developed, and constitutes a key contribution of this research. It was conceived in practical terms as PM and augments the conversation concept of the metamodel. The augmentation thus informed the name of the metamodel—Augmented Conversation and Cognitive Apprenticeship Metamodel (ACCAM). PM was implemented through integration of a novel algorithm and a user interface (i.e. a calculator). The algorithm monitors learner's learning activities on a dedicated calculator; diagnoses learner's cognitive process, interprets it and uses to aid feedback/help generation in tutoring systems. Also, as part of the CA, six teaching and learning methods—modelling, coaching, scaffolding (including fading), articulation, reflection and exploration—were proposed to achieve learning in a situated context, with learning starting from activity to abstraction. The methods constitute part of the metamodel elements (as discussed in chapter 3), and were implemented in a problem-solving context.

While the CT formed the bedrock of the metamodel in terms of the teaching and learning strategy (conversation and teachback) and KR scheme (implemented as set of rules), CA contributed in terms of the teaching and learning strategies (i.e. scaffolding, fading, etc.) and informed the learning intervention (i.e. PM) employed in the research. The latter forms the major issue investigated in this research, which attempted to unfold (or make visible) the learner's cognitive process, determine its perceived educational impact, etc. Thus, the PM implementation through an interface uniquely typified this research, and is not found in previous studies as far as we know.

Furthermore, ACCAM informed the features and functionalities that the ILABS supports. The implemented metamodel also determines the tutoring systems that can be generated (as discussed in chapter 3). ILABS was developed using Adobe Flash 4 and Action Script 3 and was desktop based. But ILATs generated from ILABS can run on a desktop as well as the Web—made possible by Adobe Flash 4. Thus, users of the ILATs could either learn offline or online as they wished. To evaluate the design approach, the ILABS and ILAT—configured via ILABS—were subjected to an extensive evaluation process. This process addressed the empirical phase of the above research objectives—which is presented below.

7.3 The Research Methodology

The methodological process, to evaluate the implemented metamodel and configured tutoring system, employed quantitative and qualitative approaches. The quantitative aspect played a dominant role, while the latter approach was utilised to gain deeper understanding of the ACCAM-based design approach.

The collection of quantitative data involved two questionnaire instruments. The first was designed to evaluate the ILABS, which was administered to authors (i.e. lecturers, the target users). The second was meant to evaluate the ILAT generated from the ILABS, and was administered to learners (i.e. students, the supposed users of the tutoring system). Some aspects of the questionnaires had some qualitative questions, but in reality, most were not completed by participants. For the qualitative data, the interview technique was employed. A series of interviews were conducted with participants in this research.

The evaluation process explained above was designed to examine the feasibility of a metamodel-based ITS authoring tool, and utilisation to produce ITSs in the problem-solving context of applied numerical domain (e.g. accounting). The possibility of generating tutoring systems that support process monitoring—practical implementation of cognitive process visibility—and model tracing was investigated. Also, the usability of the implemented metamodel (i.e. ILABS) was also investigated to determine the ease of use and other usability features. Through the ILAT (or ITS) configured, improved cognitive visibility was extensively examined. Findings from the evaluation process are presented below.

7.4 Integration and Discussion of Findings

The empirical analysis, as contained in chapters 5 and 6 of this thesis, revealed certain findings with respect to the ILABS—research objectives i and ii—and the ILAT generated from it—research objectives iii and iv. These are presented as follows:

7.4.1 Objective 1: The Metamodel and Implementation in ILABS

To conceptualise and implement a metamodel in ILABS

This research shows that a metamodel can be conceptualised, as detailed in chapter three and also highlighted above. The foregoing was demonstrated by the identification of key concepts (i.e. conversation, cognitive visibility, etc.), relevant to the context of current research. These concepts constitute the elements of ACCAM. As mentioned above, the key concepts were drawn from two learning theories—CT and CA. The research went further to prove that ACCAM can be implemented by translating the theoretical knowledge and assumptions into features implemented in ILABS (see chapter 3 - sections 3.5-3.7 for details of the implementation). The implemented metamodel (i.e. ILABS) was thereafter utilised to produce tutoring systems within the context/domain of current research. This addresses the non-empirical aspect—the design and implementation phases—of the research objective one.

On the other hand, the empirical aspect of the research objective one—the confirmation/refutation of the feasibility of the metamodel-based design approach—was undertaken through the evaluation process presented above. This process involved the analysis of authors' and learners' reactions/perceptions of the theoretical constructs (e.g. conversation, cognitive visibility, etc.,) implemented in the ILABS and tutoring systems configured. From the authors' perspective, it was confirmed that ACCAM was successfully implemented in ILABS. In addition, they noted that ILABS enabled the production of tutoring systems within the numerical problem-solving context. This was demonstrated through the features ILABS provided, which enables configuration of ILATs that contains selected the ITS authoring tool features. Also, ILATs produced demonstrated the presence of the constructs—conversation and cognitive visibility—through their behaviours/responses to learning activities. Similarly, learners further confirmed the implementation of the metamodel. They claimed that conversation and cognitive visibility concepts were embedded in the ILATs they utilised; these were observed through the ILATs' reaction to their learning activities, which enables exchange of information, interactive learning, monitors learning steps instead of goals, and provides appropriate feedbacks.

However, authors felt that the conversation aspect was marginally more pronounced as compared to cognitive visibility (mean for conversation assumption =4.67; mean for cognitive visibility assumption=4.40). Notwithstanding, their perception could be explained in the light of the design approach that enables dual-tutoring strategies. The design involved the implementation of a conversation concept as a tutoring strategy without any strict link to the ILATs' calculator. On the other hand, the cognitive visibility—implemented as process monitoring strategy—hinged on ILAT-based calculator in order to monitor learning nodes. So, it was not surprising that users made such remarks.

Also, current research noted that the calculator states (On/Off) impacted on the data entry mode. In "On" state enables implementation of PM strategy; but in the "Off" state, it only enables conversation. This thus indicates that the calculator states play significant roles in the implementation of ILABS' tutoring strategies, especially, the cognitive visibility assumption. So, non-use of the calculator disabled the process-monitoring strategy, because it was the interface responsible for the implementation of the cognitive visibility feature. The foregoing therefore proved that it was feasible to implement both process monitoring and model-tracing (i.e. conversation only) approaches within a tutoring system. This was demonstrated by the ILAT evaluated, which enhances learner's flexibility in switching the tutoring route. As well, the ILAT provides step-wise and/or goal-oriented monitoring of learning processes depending on the tutoring route chosen. Also, ILABS enables the production of ILATs without an embedded tutoring strategy. The foregoing feature enables summative evaluation of the target domain, as known in the literature (Patel, Scott & Kinshuk, 2001; Steiner & Hillemann, 2010).

Current research's empirical data further shows that a relationship exists between the assumed theoretical constructs (i.e. conversation and cognitive visibility) of the ILABS and the ILAT generated. Spearman correlation (rho) between the builder (i.e. ILABS) assumptions and the ILAT behaviour was 0.542 (normally, rho ranges between -1 to +1). Thus, the rho figure (i.e. 0.542) indicated a very strong positive correlation between the two scales measured. At least, the ILABS explained 29% of the variance in the ILAT's behaviour, which—according to Pallant (2010)—was sufficient to prove a relationship between two constructs. Thus, the research was able to prove that a metamodel can be conceptualised and implemented, thereby confirming the feasibility of a metamodel-design approach.

This research further corroborates views in the literature on the role of theoretical foundation in the development of effective tutoring systems, indicating the essentiality of theory-practice interplay (Conati & VanLehn, 2000; Harley, 2010), an approach that was achieved through ACCAM conception, followed-up with implementation and evaluation. Lastly, the research showed that it was possible to undertake ITS authoring research that has a bearing on established educational theories, unlike some authoring works (e.g. RIDES—by Munro et al., 2006—used fuzzy logic; WEAR—by Moundridou & Virvou, 2002b—no formal link to theory; Zarandi, Khademian & Minaei-Bidgoli, 2012—used fuzzy logic). Unlike SMARTIES (Hayashi, Bourdeau & Mizoguchi, 2009) which falls in the pedagogical-oriented category of ITS authoring research, ACCAM's implementation in ILABS within the performance-oriented authoring work is regarded as a significant step which attempts to answer Self's (1990b) call for the formalisation of ITS design.

The foregoing confirmed that a metamodel can be conceptualised and implemented in an ITS authoring tool, utilisable by non-programmers, to generate an unrestricted number of tutoring systems in a problem-solving context of applied numeric domains, thus addressing research objective one.

7.4.2 Objective 2—Usability of Implemented Metamodel

To assess the usability of the implemented metamodel (i.e. ILABS)

With respect to usability, the empirical data showed that the implemented metamodel (i.e. ILABS) was easy to use and usable. However, at the first attempt of usage, ILABS might require introductory human support or a user manual. As a result, this research considered incorporating help facilities in future versions to eliminate or reduce the need for initial human support and enhance the learning curve. Despite that, current research noted that ILABS did not require extensive training, and familiarisation tends to improve its usage. Authors confirmed that the functionalities embedded in the ILABS achieved their purpose. However, they suggested the need to incorporate a recovery mechanism in ILABS, to prevent any occasional freezes that might occur due to a bug which is normal with any software. ILABS provided facilities to open existing data files for modification/reconfiguration and/or extension of ILATs. Notwithstanding, authors were of the view that the ILABS should provide import and conversion facilities. They envisaged such features would enable files created in other applications to be usable within ILABS, thereby enhancing the ITS authoring tool's usability. The suggested functionalities were earmarked for future versions of the ILABS.

The empirical data revealed that ILABS requires setting attributes of the data objects embedded in ILATs produced. ILABS also requires the re-generation of existing ILAT after modification to effect changes, as part of the implementation procedures. Also, the empirical data suggested that production was achieved within a short span of time. However, production turnaround time was predicated on topic/problem complexity. The foregoing is understandable since the nature and complexity of a domain topic determines the volume of data objects and interface structure configured in a tutoring system. Notwithstanding, since ILABS is a prototype, improvement to its usability is envisaged in future work.

This research showed that ILABS is unrestrictive in terms of the number of ILATs and their variants that can be produced. An exception is that ILABS may not produce tutoring systems for complex topics/problems due to its limited arithmetic operators and features. On the latter remark, provision was made to enable importation of customised routines/functions into ILABS, which can be used to handle complex scenarios. However, the importation feature could not be finalised in the course of this work, and does not—in any way—limit the implemented metamodel. In addition, authors acknowledged that programming skill was not required to use the ILABS. However, basic computer literacy and prior knowledge of the implementation domain were required to use the ILABS effectively and configure useful ILATs. In fact, authors remarked that ILABS was a successful implementation of ACCAM having achieved its purpose. That is, ILABS was a typical example of a formal (theory-based) ITS authoring design and enables configuration of tutoring systems by non-programmers.

Although this research acknowledged the above issues, past ITS authoring works are not exempted (e.g. SMARTIES—Hayashi, Bourdeau & Mizoguchi, 2009; WEAR—Moundridou & Virvou, 2002a; REDEEM—Ainsworth & Fleming, 2006). The foregoing studies had their successes, as well as their weaknesses, which should be addressed. As an example, REDEEM (Ainsworth & Fleming, 2006)—an ITS authoring tool—was subjected to evaluation to determine its extent of usability by non-programmers to construct the declarative aspect of knowledge. It provided a simple interface that enables teachers who may not be familiar with computer technology to create learning materials easily. The success of REDEEM was attributed to the extent the authoring tool was usable by its intended users (non-programmers). Nevertheless, they acknowledged the need to improve the design of REDEEM in order to enhance its functionality. Moundridou & Virvou (2002b) carried out an evaluation of WEAR—an ITS authoring tool for algebra-related domains. They also claimed the need for the enrichment of authors'/instructors' role through provision of relevant information during the ITS construction using the WEAR authoring tool. Providing help features in ILABS, as identified in this research, could be said to be similar to the provision of relevant information in WEAR.

REDEEM was based on meta-strategy that comprises conditions and strategy descriptions specification by way of parameter-setting (Hayashi, Bourdeau & Mizoguchi, 2009), which enhances its usability by non-programmers. However, REDEEM and WEAR designs could not be linked to any explicit theoretical knowledge, since they were not underpinned by any educational theory. Users need to understand the learning/instructional theories to configure learning materials (Hayashi, Bourdeau & Mizoguchi, 2009). This places an extra burden on users. This is not the case with ILABS, since it provides pedagogical focus through the constructivist theories (CT & CA) that underpin it. On the other hand, SMARTIES was linked to theories and utilised AI agents to infer intelligence. In contrast, ILABS is based on ACCAM and AI-neutral; it employs a calculator interface in configured ILATS to capture the cognitive process of a learner, thereby eliminating the inaccuracy that may result from extensive AI inference techniques. While SMARTIES was based on ontology engineering, which has been acknowledged to increase the burden of users due to its complexity (Hayashi, Bourdeau & Mizoguchi, 2009), ILABS was based on ACCAM and has been evaluated to be usable by non-programmers.

Therefore, issues identified with respect to the ILABS, do not in any way invalidate the success achieved. Instead, they constitute part of the ILABS development process, and usually such issues come up in evaluation studies of this nature. Furthermore, they indicate areas that can be worked upon to improve the standard of the ILABS and enhance its usability. Also, while REDEEM and SMARTIES were implemented for the declarative aspect of knowledge, the ILABS discussed in this thesis was a success story in the procedural aspect of knowledge and in applied numerical disciplines. Thus, the foregoing further confirms the feasibility and usability potentials of the ACCAM-based design approach.

7.4.3 Objective 3—Using PM for Improved Cognitive Visibility

To evaluate the use of process monitoring to increase visibility of the cognitive process of a learner

This research showed that the learner's cognitive process can be made visible. Learners confirmed that the process-monitoring algorithm developed—to implement the cognitive visibility concept—as part of the ACCAM's implementation, worked as designed. The algorithm constitutes one of the selectable and optional tutoring strategy routes (process monitoring, model-tracing and non-tutoring) that ILABS provided.

In the ILAT, the strategy—that is, PM—was demonstrated through tutoring system's behaviour. PM enforces learning via the ILAT's calculator, being the key implementation medium. The calculator constitutes the ILAT's interface that provides learning inputs that the PM algorithm monitors and interprets. As part of the behaviour PM exhibited in the tutoring system, the calculator captures the learning process in a step-wise pattern, mapping learning nodes as the learning progresses. Consequently, through the step-wise pattern mapping, this research revealed that learners' cognitive process was made visible. This was only possible when learning took place via the ILAT's calculator.

The empirical data confirmed that learner's misconception (including missing conception—jumping solution step[s]) was identified by the tutoring system. Evaluators (i.e. learners) observed two misconception points, step and goal levels. These points depended on the tutoring route taken (process monitoring or model-tracing strategy respectively). Hence, the empirical data suggested that—through cognitive visibility—learners' misconception can be tracked which enhanced the diagnosis of their learning process. Consequently, guidance can be provided to enhance learning.

While the foregoing further confirms enhanced learning diagnosis—through cognitive visibility—similar to closely-related concepts (e.g. cognitive mapping, plan recognition, etc.) in previous studies, the PM approach adopted differs from other works in the literature. For instance, previous implementations of cognitive-related issues utilised AI techniques, such as fuzzy logic, BNs (Conati & VanLehn, 1996; Conati, Gertner & VanLehn, 2002; Woolf, 2009; Chieu et al., 2010), etc, whereas the implementation of cognitive visibility—in this research—stems from a theoretical foundation (that is, ACCAM), and uses an algorithm that is rule-based to interpret learners' inputs via a calculator interface. This capability was then integrated as part of features ILABS provides. Results from the empirical data indicated that the current research's implementation approach of cognitive visibility worked. Thus, it provides insights into other ways of achieving cognitive visibility and intelligence in ITS. This approach also shows that the implementation can be created as a feature of an ITS authoring tool.

7.4.4 Objective 4—Impact of PM on Feedbacks and Perceived Learning Effectiveness

To determine the perception of target users regarding the impact of process monitoring/cognitive visibility on feedback and learning effectiveness.

The users of the ILAT, that is learners, noted that process monitoring impacted feedback. ILAT provides feedback in the form of reward or guidance (as a response to identified misconception discussed above). When the PM route was chosen by learners, ILAT was able to provide immediate feedback. However, when the model-tracing route was taken, ILAT provided delayed feedback. Immediate feedbacks were given as soon as misconceptions were identified at step-level, while delayed feedbacks were given at goal level. Accordingly, the embedded ILAT's feedback generator relates feedback to problem states. Note that the feedback generator algorithm was integrated with the tutoring strategies provided by ILABS. The empirical data thus suggested that PM enables the provision of timely and relevant feedback whenever misconception is identified. This research proved that there was an established link between the process monitoring and feedback generator algorithms that were developed, and this link worked as envisaged or designed. The current research conclusively shows that process monitoring, which was instrumental to improved cognitive visibility and identification of learning misconceptions/missing conceptions, actually enhanced the generation of relevant and timely feedback.

Similarly, this research further provides insight into the learning effectiveness of cognitive visibility. Learners indicated that cognitive visibility—as implemented via process monitoring—enhance learning. The perceived learning effectiveness of PM was predicated on the prompt identification of misconception/missing conception, and provision of appropriate guidance by ILAT. Aleven et al. (2009) showed that the provision of guidance, especially step-by-step guidance, was effective in enhancing learning. Likewise, Melis (2005) and Shute (2008) stressed that feedback plays a significant role in the achievement of effective learning. So, the foregoing suggests that the current implementation, which provides feedback in response to identified misconception/missing conception at step levels, potentially provides effective learning experience. This stance was reflected in learners' perception, which shows that PM effectively impacted learning. Furthermore, learners' belief regarding the likely learning effectiveness of cognitive visibility, if implemented in a tutoring system, tallied with their perception evaluation.

However, learners felt that cognitive visibility—through PM—was more useful at the early stage of learning due to the prompt feedbacks/guidance they received; it was helpful in shaping learning at this stage. At an advanced stage, frequent feedback appeared to interfere with smooth learning, because it prevents reflection. Learners want to think through the problem at hand and change their actions before committing their answers. Nonetheless, ILAT makes provision such that PM can be turned off/on as required, thus allowing learners to take a different tutoring route. Qualitative data suggests that some learners turn off PM at advanced stage of learning, thus hindering learning diagnosis when required. Consequently, future work intends to extend the implementation such that the cognitive process will be captured at all times once PM is "ON", but feedback will only be given when queried. Such extended implementation will prevent learning interference and enhance diagnosis at any stage of learning (i.e. early or advance stages).

Overall, this research proves that the current PM implementation worked, and is educationally viable. Thus, the research results show that cognitive visibility (as a construct and teaching strategy), aligns with ACCAM's assumptions—that if the cognitive process is made visible, learning will be effective.

7.5 Discussion of Other Findings

Apart from the above, some other findings emerged from the empirical study. These are discussed below.

7.5.1 Learning Process & Elements

The current research revealed that the implemented metamodel (i.e. ILABS) enabled production of ILATs that promote a learning process, which is tagged ACCAM-LP—that is ACCAM Learning Process. The empirical data indicated the components of the ACCAM-LP to be bi-directional communication and interactive learning (signposting conversation), step-wise monitoring (signposting cognitive visibility) and/or goal-oriented monitoring, misconception (including missing conception), feedback, reflection, remediation and evaluation. This research also links motivation to learning. According to authors, most learners tend to use learning tools to fulfil academic requirements. As a result, the empirical data appears to suggest the inclusion of motivation features in tutoring systems. Also, it appears to justify the need to include motivation in the elements of a learning process for effective learning to take place.

While noting the above, Buckler (1996) presented a learning process model for business organisations. The latter acknowledged that the process model was made of components that facilitated learning within a business organisation. This is comparable to the ACCAM-LP identified in this research, which addressed learning within an educational environment. However, the success of Buckler's learning process model was predicated on the quality of leadership provided by managers and team leaders. It can therefore be inferred that the success of any system would depend on the level of integration of its components and drivers. Accordingly, this research showed that the ACCAM-LP was effective in enhancing learning. Its effectiveness was due to the twin elements of conversation and cognitive visibility, which enhanced the emergence of other elements of the learning process. However, the success of ACCAM-LP will depend on the integration of ILATs in the curriculum, as part of the teaching and learning support tools. If authors (i.e. lecturers) fail to adopt ILATs, learners may be unwilling to use ILATs for the tutorial aspect of their learning except those that are positively disposed. Thus, authors are key drivers to the success of the ACCAM-LP since their stance with respect to ILATs can positively or negatively impact learners' motivation towards the tutoring systems.

Also, the ACCAM-LP demonstrated certain learning characteristics that are comparable to the attributes of meaningful learning—i.e. active (or constructive), intentional (or reflective), authentic (or contextualised), and cooperative (or collaborative/conversational) learning (as cited in Pongsuwan et al., 2011). These attributes were exhibited by the ACCAM-LP as follows: [i] learners are involved in knowledge construction (i.e. active learners construct knowledge via problem solving—constructive); [ii] it encourages reflection via misconceptions and remediation (i.e. intentional or reflective); [iii] learning takes place in a problem-solving context (i.e. contextualised); and [iv] it promotes learning via conversation (i.e. cooperative). The foregoing thereafter enhanced cognitive visibility of learners. Thus, the ACCAM-LP elements impacted upon each other, which consequently contributed to effective learning that was achieved via the ILAT.

7.5.2 Learning Behaviour

Authors noted that non-reconfigurable tutoring systems (i.e. ITSs that cannot be personalised), tend to promote surface-learning behaviours (such as memorising and regurgitating solution steps of given problems/topics), instead of learning to understand topic concepts and their relationships. However, the introduction of ILABS was perceived to eliminate, or curbs, such surface-learning tendencies, since it supports features that enable reconfiguration and personalisation of tutoring systems. Inactive learning attitude was also associated with lack of motivation or interest in learning. The foregoing further strengthens the role of motivation in promoting positive and active learning.

Also, authors claimed that learners might prefer to use their own calculator rather than the ILAT based calculator. However, it should be noted that the use of the calculator interface to implement PM was deliberately chosen to evaluate the learning effectiveness of the interface-based PM. Despite that, this research envisaged that an alternative learning route, which is not linked to the calculator interface, should be devised. The provision of multiple tutoring strategies in ILABS thus addresses the foregoing situation. Learners with preference for their own calculator can opt for model-tracing learning route rather than PM, if configured in the ILAT being used. Thus, the support of multiple strategies enhances the flexibility of the implementation to accommodate different learning behaviours.

7.6 Implications for Research and Practice

The research findings are likely to impact research in the field of ITS/Authoring in the following ways:

- The outcome of this research, when considered with previous arguments in the literature (Self, 1990b; Harley, 2010), tends to strengthen the need for theoretical foundation for ITS/Authoring tools. Moreover, if they are educational tools, they should be driven by educational objectives.
- This research calls for change in the approach adopted in the construction of ITS/Authoring tools. Features of such tools should be determined by their underpinning educational theories. This would enable the provision of

ITS/Authoring tools that have educational values, and can thus be evaluated based on their educational premise.

- Although the research revealed that the ILABS was usable, it identified the need for help facilities. This implies that ITS/Authoring development should incorporate help features in their design, since it could enhance understanding/utilisation of concepts or functionalities built into them.
- Also, the development of ITS authoring tools should take into consideration the learning behaviour of students, as this will impact the usefulness of any tutoring system configured. Issues of motivation should be considered very important when configuring any tutoring system, since motivation seems to be a key driver of learning.

7.7 Contributions of the Research

Current research efforts in the fields of ITS and authoring focused on bringing more reliable and effective learning tools to the classroom. These developments can be attributed to the migration of researchers in AI and cognitive science to education. The growing research efforts in the fields of ITS/Authoring are also attributable to: the growing educational space, the need to support the traditional approach to pedagogy, the growing number of students requiring learning attention, the need to provide virtual education, and other educational needs.

This research is not an exception; rather, it attempts to fill some gaps—earlier mentioned in chapter two—in ITS/Authoring research, that were found to be crucial in order to enhance students' learning experience. Current research also attempted to provide a platform through which more effective learning tools could be constructed for applied numerical disciplines (e.g. accounting and finance) in order to address some learning difficulties generally associated with such disciplines.

In accordance with above, this research contributes to knowledge and practice in the field of ITS/Authoring as captured in figure 7.1 below. The key contributions are further explained below.

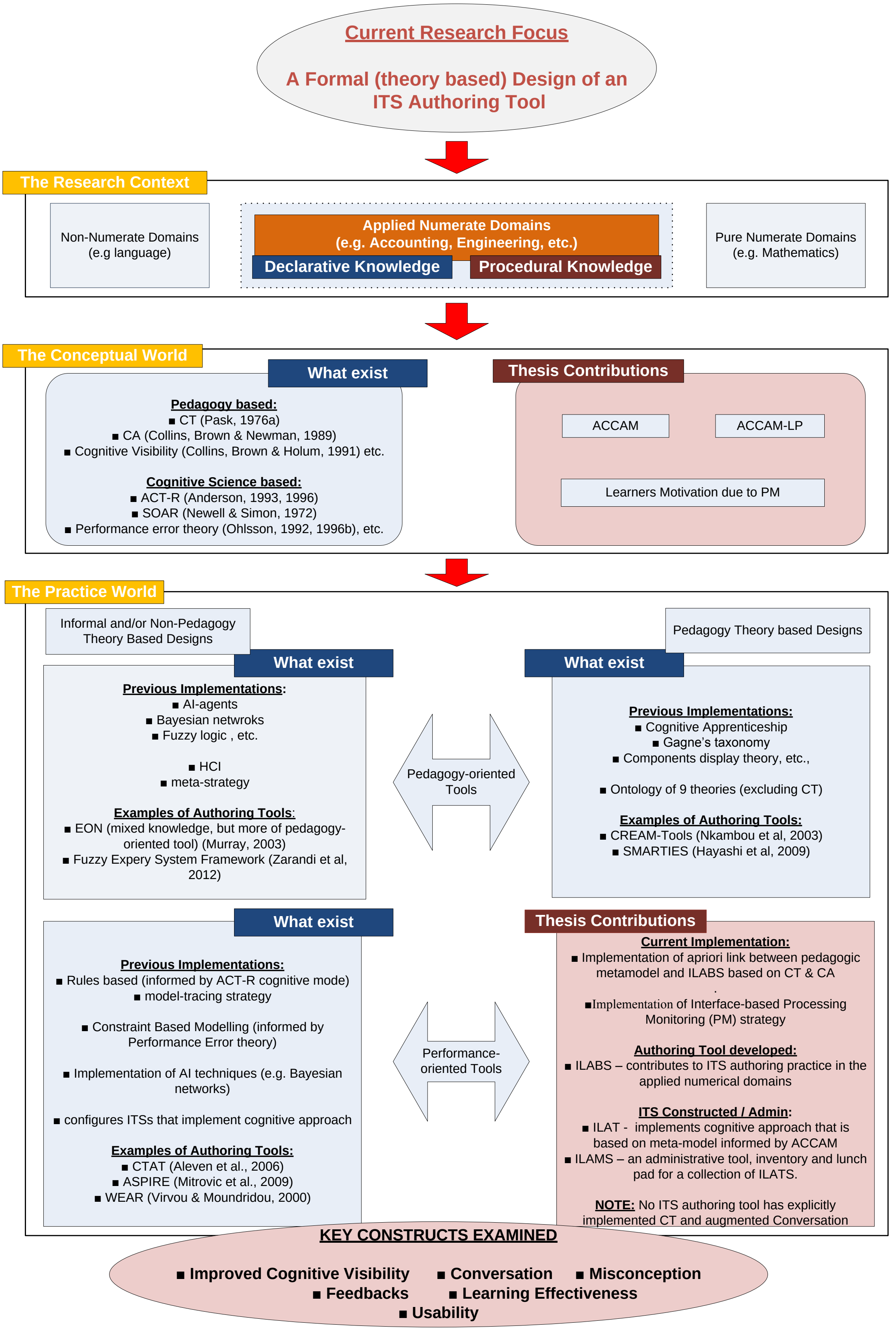


Figure 7. 1: Contribution of the Current Research

7.7.1 Conceptualisation of ACCAM

The conception of an Augmented Conversation and Cognitive Apprenticeship Metamodel (ACCAM), from two specific pedagogical theories—Conversation Theory (CT) and Cognitive Apprenticeship (CA), was undertaken in this research. Although, CT and CA have been used in previous ITS work, this work is the first attempt to augment conversation within a cognitive apprenticeship framework and conceptualised as ACCAM. It is also the first attempt to shape the design of an ITS authoring tool using ACCAM that is based on the foregoing specific pedagogical theories. The choice of these theories arose from the suitability of the application of their conceptual frameworks—conversation and cognitive apprenticeship—within the numerical problem-solving context of the applied numerical domains in which this research is undertaken. This is so, since it has been established that learning-applied numerical domains involve a lot of cognitive tasks, and conversation is considered a suitable medium through which a domain expert can exchange knowledge or information with a learner (Patel, Scott & Kinshuk, 2001).

The augmentation aspect of ACCAM constitutes a principal element of the theoretical platform for a formal design and construction of educational tools which can be used to further enhance learning by improving the visibility of the cognitive process of a learner. Improved cognitive visibility is required, since it has been theoretically argued that if the cognitive process of a learner can be made open, the domain expert (or master) will be positioned to provide reliable and useful guidance to a learner (or novice) in order to enhance knowledge/skill construction (Collins, Brown & Holum, 1991). This is so since a domain expert may need to diagnose the misconception of a learner, in order to determine the guidance to provide during knowledge construction.

Consequent to the above, ACCAM is considered a novel conceptual step towards achieving an open learning space that promotes effective learning. Also, the metamodel provides the conceptual basis for the investigation of improved visibility of learners' cognitive process in order to determine the effectiveness of the ACCAM-based design approach. Therefore, this work is an original contribution to knowledge being the first attempt to abstract a metamodel from two constructivist theories, involving augmentation of learning conversations within a cognitive apprenticeship framework. By so doing, it addressed the need for a formal (theory based) approach to the design of ITS/Authoring tools, as argued/identified in chapter 2 of this thesis.

7.7.2 Cognitive Visibility through Process Monitoring (PM)

This research proved that the visibility of a learner's cognitive process can be improved through the augmentation of learning conversations using PM via an interface (i.e. a calculator). Previous studies had used other approaches such as BNs (Conati & VanLehn, 1996) and fuzzy logic (Zarandi, Khademian & Minaei-Bidgoli, 2012) to investigate other closely-related concepts (e.g. cognitive mapping, plan recognition, etc.). The current approach took advantage of a practice in the applied numerical domain—accounting and finance—which benefited from the extensive use of a calculator during problem-solving situations. The evaluation aspect of this research shows that the implementation of PM via an interface was successful. It shows that PM improved learning conversations and revealed the cognitive nodes of learners during the learning process. Also, PM improved the detection of learners' misconceptions, provided feedback generation and was effective in enhancing learning.

Based on the above, this research was able to provide a conceptual understanding of how cognitive visibility—implemented as PM—impacts learning, the detection of misconception and the generation of feedback in a technology-based learning environment. This aided the comprehension of the role of cognitive visibility in the learning of procedural knowledge of applied numerical domains (e.g. accounting and finance). Therefore, the conceptualisation and practical implementation of cognitive visibility via an interface (i.e. a calculator) is regarded as a unique and significant contribution, which no previous studies have undertaken. Also, it demonstrates an alternative approach to the implementation of intelligence in tutoring systems outside the standard AI techniques predominantly utilised in the field. The current implementation of intelligence—through PM and an interface—has an advantage over standard AI techniques that mainly rely on inference, which may occasionally fail. Therefore, this aspect of the current research was also an original contribution, since no such conception/implementation of cognitive visibility and intelligence has been undertaken in the past.

7.7.3 ITS/Authoring Practice

The implementation of ACCAM contributes to practice by establishing a formal link between theory and practice which yielded the ILABS—a practical implementation of the metamodel. This was undertaken by implementing ACCAM in ILABS, which shapes and determines the features that constitute the ILABS and the ILAT constructed. Thus, based on Murray's (1999, 2003a) classification, this work stands to be the first

ITS authoring work which is underpinned by two specific pedagogical theories—CT and CA through ACCAM—and explicitly falls within the performance-oriented category, unlike SMARTIES that falls within the pedagogy-oriented category. A performance-oriented authoring tool focuses on providing a rich learning environment which supports learning by practice and receiving feedback. On the other hand, pedagogy-oriented tools focus on how to sequence and teach canned domain content. Thus, the implementation of ACCAM in ILABS has the advantage of providing a pedagogical focus that can aid a curriculum designer's decision-making, since it is underpinned by constructivist theories (CT and CA). This is in contrast to SMARTIES that attempts to infer author's pedagogy which may not be accurately detected.

Also, the current research acknowledged that research is undertaken in context and each discipline works within its own frame of reference (Luckin, 2010). So, this research was undertaken and contributes to the numerical problem-solving context of the procedural aspects of the applied numerical domains. It provides an authoring platform to generate tutoring systems that enable knowledge construction through learning by doing. As a result, this work enhances learning of the procedural knowledge—an aspect of knowledge that appears to distinguish a domain expert from a novice (Patel, Scott & Kinshuk, 2001). In contrast to researches predominantly conducted in pure numerical domains (e.g. mathematics), this work adds to authoring research in the applied numerical domain—specifically, accounting—that have been rarely patronised. ILABS enables the construction of ILATs for the numerate aspect of accounting discipline. By so doing, it provides a cost-effective means of constructing ITSs that is based on formal (theory based) design approach.

7.8 Limitations of the Research

Despite the success achieved, the research could have taken care of some issues, but could not do so, due to some research constraints, such as time, cost, among others. These issues are highlighted below.

- The research could have been tested in the numerical problem-solving context of other numerical disciplines, outside accounting and finance, in order to enhance applicability of the ILABS in other domains. This was not achieved due to certain requirements, including inputs from alternative domains experts. Time and accessibility to such experts were major limiting factors. Significant amounts of time are required to contact and secure the audience of appropriate domain experts. Their expertise would actually be needed, if the research were extended to other domains. Inability to achieve this, therefore, limits the extent of generalisation to other numerical domains. This could have enabled the investigation of the capability of ILABS to generate usable tutoring systems in other disciplines. Moreover, it could have enabled examination of process monitoring (i.e. making cognitive process visible) in other numerical domains.

- It could be observed that, for research question one—that addresses objective one and two above, there was a large difference between the quantitative and qualitative samples. Moreover, both approaches addressed the same research question/propositions, and the qualitative samples were drawn from those who participated in the quantitative aspect of the research. Nevertheless, it does not—in any way—invalidate the conclusion reached, since the findings from both approaches did not contradict each other. Instead, the qualitative findings provided deeper insights into findings that were revealed. Such research situations have also been acknowledged in the literature (see Creswell & Plano Clark, 2007, p.120).
- Experimentation has occupied many evaluation studies in the literature. It has been noted to be the most effective way to measure the learning effectiveness of an educational intervention (e.g. PM). However the research was constrained; specifically, getting students to participate in this type of evaluation was problematic, accompanied by its cost and time implications. Thus experimentation on the learning effectiveness of the PM strategy employed in the research could not be undertaken.
- Also, to achieve the best measure of learning effectiveness, experimentation should be conducted in the real world of use (i.e. in classroom as part of the curriculum). However, it is unethical to prevent a set of learners the use of ILATs with PM while allowing another set in a real world of testing, thus limiting what can be achieved.
- The evaluation could not measure the affect aspect of the ILABS and the ILATs generated from it. This was considered an important educational component that should have been considered, since the research revealed that motivation enhances learning. A research technique, such as observation, could have been explored to record the affect aspect of the implemented metamodel and modules generated from it. This could have extended the research, enabling coverage of

the two key aspects that constitute educational impact (i.e. affect and learning effectiveness) as identified in Mark & Greer (1993).

- Also, a multi-evaluation strategy incorporating several strategies in a layered manner as suggested in some studies (see Brusilovsky, Karagiannidis & Sampson, 2001; Brusilovsky, Farzan & Ahn, 2006; Paramythis, Weibelzahl & Masthoff, 2010) could have been explored, and would have benefited this research. Such strategies could have included Action Research to enrich the qualitative aspect of the work. The qualitative strategy could have been extended to include the ITS evaluation as well. This could provide a richer data set that the research could analyse to reveal knowledge that could form the bedrock for new thinking in the ITS/Authoring field.

7.9 Future Direction

The research discussed in this thesis extensively dealt with testing some theoretical assumptions through a metamodel-based ILABS and the construction of some Intelligent Tutors. This provided the ground to evaluate a formal (theory based) design approach. The ILABS and generated tutoring systems were subjected to evaluation, involving quantitative and qualitative analyses, to confirm or refute the theoretical assumptions that underlie the design approach. Despite this, there are other research openings that could still be subjected to further investigation within the current work. Some of these openings, as a result of the thesis contribution to knowledge, and due to observed limitations of the research, are discussed in the following paragraphs.

It was observed that consideration could be given to affect issues in the design and development of the ILABS. This area, which includes users' attitude and emotion towards a computer-based tool, could be investigated. Although there are some works on the theme "affect" in relation to ITS, none could be seen with respect to ITS authoring tools. In this research, the design and development of the ILABS took a simplistic approach, which could make or mar reuse of the tool. The approach may not be so attractive to some users, who are interested in a "fancy" interface. The impact of this approach in terms of the reuse value is an issue that is open for exploration. This may contribute to future improvement of the interface design, as well as enable the formation of theoretical grounds through the testing of "affect" theories.

The metamodel implemented has other assumptions, which were not tested in the evaluation carried out. Although practically, these assumptions could be observed in the functionalities made available while using the tool, an empirical confirmation that could prove their embedment in the ILABS and ILATs generated would be necessary. Equally, their educational impact with respect to their use in this research would either strengthen or weaken their consideration, which may inform whether to retain or remove them from the theoretical assumptions.

Just as it was identified under the limitations of the research, the evaluation methodology adopted could be extended. Evaluation could be carried out in a layered manner involving multiple research strategies and techniques in a multi-institutional evaluation, where feasible. This would provide very rich data sets that could be worked upon to really test the metamodel-based approach, determine the adherence to the theoretical assumptions of underlying theories and ascertain their educational impact in its totality. This would also help inform the design and development strategy that could be utilised in future research of this nature.

Another viable aspect of this research that could be investigated relates to the inner workings of the ILABS that was developed. Current research did not evaluate the algorithms that were developed. It could be necessary to carry out robust testing of these algorithms to authenticate their individual and integrated functionality within the ILABS and ILATs generated, in order to match them with the design goals.

Also, the research revealed that the ILABS should be extended to generate: mobile-based tutors (to widen accessibility, instead of restricting learning to computers only) and game-like tutors (to enhance motivation, since students love games). This particular aspect confirms the importance of motivation in learning. The mentioned areas, as revealed in the empirical studies, could be explored in future.

References

Ainsworth, S. & Fleming, P. (2006). Evaluating authoring tools for teachers as instructional designers. Computers in human behavior, 22(1), 131–148.

Ainsworth, S. & Grimshaw, S. (2004). Evaluating the REDEEM authoring tool: can teachers create effective learning environments? International journal of artificial intelligence in education, 14, 279-312.

Ainsworth, S., Major, N., Grimshaw, S., Hayes, M., Underwood, J., Williams, B. & Wood, D. (2003). REDEEM: Simple intelligent tutoring systems from usable tools. In T. Murray, S. Ainsworth, & S. Blessing (Eds.), Authoring Tools for Advance Technology Learning Environments (pp. 205-232).

Akilli, G.K. (2005). User Satisfaction Evaluation of an Educational Website. The Turkish Online Journal of Educational Technology-TOJET, 4(1), 85–92.

Akin, O. (1986). Psychology of Architectural Design. Pion Ltd., London.

Akin, O. (2008). Frames of Reference in Architectural Design : Analyzing the Hyper-Acclamation (A-h-a-!). School of Architecture, Paper 25, 1-20. Retrieved from http://repository.cmu.edu/architecture/25

Aleven, V. & Koedinger, K.R. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. Cognitive Science, 26(2), 147-179.

Aleven, V., Koedinger, K.R. & Cross, K. (1999). Tutoring Answer Explanation Fosters Learning with Understanding. In Artificial Intelligence in Education, S. P. Lajoie and M. Vivet (eds.), 199–206. Amsterdam: IOS.

Aleven, V., McLaren, B., Roll, I. & Koedinger, K. (2004). Toward Tutoring Help Seeking (Applying Cognitive Modeling to Meta-cognitive Skills). In J. C. Lester, R. M. Vicario, & F. Paraguaçu (Eds.), Proceedings of Seventh International Conference on Intelligent Tutoring Systems, ITS 2004. (pp. 227–239). SPRINGER-VERLAG. Retrieved from http://repository.cmu.edu/cgi/viewcontent.cgi?article=1155&context=hcii

Aleven, V., McLaren, B., Roll, I. & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. International Journal of Artificial Intelligence in Education, 16(2), 101–128. IOS Press.

Aleven, V., Sewall, J., McLaren, B.M. & Koedinger, K.R. (2006a). Rapid Authoring of Intelligent Tutors for Real-World and Experimental Use. In K. Kinshuk, R. Koper, P. Kommers, P. Kirschner, D. G. Sampson, W. Didderen, & D. (Eds.) (Eds.), Sixth IEEE International Conference on Advanced Learning Technologies (ICALT’06) (pp. 847-851). Los Alamitos, CA: IEEE Computer Society.

Aleven, V., McLaren, B., Sewall, J. & Koedinger, K. (2006b). The cognitive tutor authoring tools (ctat): Preliminary evaluation of efficiency gains. In M. Ikeda, K. D. Ashley, & T. W. Chan (Eds.), Proceedings of the 8th International Conference on Intelligent Tutoring Systems (pp. 61–70). Berlin: Springer Verlag: Springer. Retrieved from http://www.springerlink.com/index/K037578X2Q825156.pdf

Aleven, V., Mclaren, B. M., Sewall, J. & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. International Journal of Artificial Intelligence in Education, 19(2), 105–154. IOS Press.

Aleven, V., Roll, I., McLaren, B., Ryu, E. J. & Koedinger, K. (2005). An architecture to combine meta-cognitive and cognitive tutoring: Pilot testing the Help Tutor. Proceeding of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology (pp. 17–24). IOS Press. Retrieved from http://portal.acm.org/citation.cfm?id=1562533

Ali, A. M. & Yusof, H. (2011). Quality in Qualitative Studies: The Case of Validity, Reliability and Generalizability. Issues in Social and Environmental Accounting, 5(1/2), 25-64.

Alpert, S.R., Singley, M. & Carroll, J.M. (1999). Multiple instructional agents in an intelligent tutoring system. In AIED'99 Workshop on Animated and Personified Pedagogical Agents, 10–17. Le Mans (pp. 1-8). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.85.3588

Alseddiqi, M. & Mishra, R. (2010). A Diagnostic Study On The Teaching And Learning Styles In Engineering Education. Computing and Engineering Researchers' Conference, University of Huddersfield. (pp. 105-109).

Altman, D., Burton, N., Cuthill, I., Festing, M., Hutton, J. & Playle, L. (2006). Why do a pilot study? National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs, 2006). Retrieved from http://www.nc3rs.org.uk/downloaddoc.asp?id=400

Alves, P., Pires, J. A. & Amaral, L. (2009). The role of learning styles in intelligent tutoring systems. International Conference on Computer Supported Education (Vol. I, pp. 315-320). Lisboa, 2009: vol I; Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+role+of+learning+styles+in+intelligent+tutoring+systems#0

Amoia, M., Gardent, C. & Perez-Beltrachini, L. (2011). A serious game for second language acquisition. Proceedings of the 3rd International Conference on Computer Supported Education (CSEDU 2011). Noordwijkerhout, The Netherlands. Retrieved from http://talc.loria.fr/IMG/pdf/CSEDU2011nancy.pdf

Anderson, J.R. (1990). The Adaptive Character of Thought. Hillsfdale, NJ:Erlbaum.

Anderson, J.R. (1993a). Rules of The Mind. New Jessey: Lawrence Erlbaum Associates, Inc.

Anderson, J.R, Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C. & Qin, Y. (2004). An integrated theory of the mind. Psychological review, 111(4), 1036-60.

Anderson, J.R, Boyle, C., Farrell, R. & Reiser, B.J. (1984). Cognitive principles in the design of computer tutors. In Proceedings of the Sixth Annual Conference of the Cognitive Science Society (pp. 2-9). Boulder, CO: University of Colorado.

Anderson, J.R, Boyle, C., Farrell, R. & Reiser, B.J. (1987). Cognitive principles in the design of computer tutors. In P. Morris (Ed.), Modelling cognition (pp. 93-133). John Wiley & Sons Ltd.

Anderson, J.R., Boyle, C.F. & Yost, G. (1985). The geometry tutor. Proceedings of the Ninth International Joint conference on Artificial intelligence (Vol. 1, pp. 1–7).

Anderson, J.R., Corbett, A.T., Koedinger, K.R. & Pelletier, R. (1995). Cognitive tutors: Lessons learned. Journal of the Learning Sciences, 4(2), 167–207. Routledge.

Anderson, J.R, Boyle, C.F. & Reiser, B. J. (1985). IntelligentTutoringSystems. Science, 228, 456-468.

Anderson, J.R. & Lebiere, C. (1998). The Atomic Components of Thought. Lawrence Erlbaum Associates. (Online version: abstracts). Accessed June 14, 2011 from: http://act-r.psy.cmu.edu/book/

Arroyo, I., Beck, J.E., Beal, C.R., Wing, R. & Woolf, B.P. (2001). Analyzing students' response to help provision in an elementary mathematics intelligent tutoring system. Papers of the AIED-2001 Workshop on Help Provision and Help Seeking in Interactive Learning Environments (pp. 34–46). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.25.2520&rep=rep1&type=pdf

Arroyo, I., Royer, J. & Woolf, B.P. (2011). Using an Intelligent Tutor and Math Fluency Training to Improve Math Performance. International Journal of Artificial Intelligence in Education, 21, 135-152.

Arruabarrena, R., Pérez, T., López-Cuadrado, J., Gutiérrez, J. & Vadillo, J. (2002). On evaluating adaptive systems for education. In Proc. AH2002, 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (Vol. 2347, pp. 363–367). Malaga: Springer-Verlag. Retrieved from http://www.springerlink.com/index/hlj91ajgcbmdcfql.pdf

Arruarte, A., Ferrero, B., Fernandez-Castro, I., Urretavizcaya, M., Alvarez, A. & Greer, J. (2003). The IRIS Authoring Tool. In T. Murray, S. Ainsworth, & S. Blessing (Eds.), Authoring Tools for Advance Technology Learning Environments (pp. 233-267).

Auld, G.W., Diker, A., Bock, M.A., Boushey, C.J., Bruhn, C.M., Cluskey, M., Edlefsen, M., et al. (2007). Development of a decision tree to determine appropriateness of NVivo in analyzing qualitative data sets. Journal of nutrition, 39(Figure 1), 37-47.

Baker, R., Goldstein, A. & Heffernan, N. (2011). Detecting Learning Moment-by-Moment. International Journal of Artificial Intelligence in Education, 21, 5-25.

Baker, R.S., Corbett, A.T. & Koedinger, K.R. (2004). Detecting student misuse of intelligent tutoring systems. Intelligent tutoring systems (pp. 54–76). Springer.

Barker, P. (2008). Re-evaluating a model of learning design. Innovations in Education and Teaching International, 45(2), 127-141.

Barnes, T. & Stamper, J. (2010). Automatic Hint Generation for Logic Proof Tutoring Using Historical Data. Educational Technology & Society, 13(1), 3-12.

Barrow, D., Mitrovic, A., Ohlsson, S., Grimley, M. et al. (2008). Assessing the impact of positive feedback in constraint-based ITSs. In:Woolf, B. (ed.) Proceedings of 9th International Conference on ITS 2008, LCNS 5091, pp. 250–259. Springer-Verlag, Heidelberg.

Baylari, A. & Montazer, G.A. (2009). Design a personalized e-learning system based on item response theory and artificial neural network approach. Expert Systems with Applications, 36(4), 8013-8021. Elsevier Ltd.

Bazeley, P. (2009). Analysis Qualitative Data: More Than Identifying Themes. Malaysian Journal of Qualitative Research, 2, 6-22.

Bazeley, P. (2004). Issues in mixing qualitative and quantitative approaches to research. In R. Buber, J. Gadner, & L. Richards (Eds.), Applying qualitative methods to marketing management research (pp. 141-156). Hampshire, United Kingdom: Palgrave Macmillan.

Beaumont, R. (2009). Experimental Design: A set of notes suitable for seminar use. Design. Retrieved from http://www.fhi.rcsed.ac.uk/rbeaumont/virtualclassroom/chap16/s1/sembk2.pdf

Beck, J., Stern, M. & Haugsjaa, E. (1996). Applications of AI in Education. Crossroads, 3(1), 11–15. ACM.

Ben Ammar, M., Neji, M., Alimi, A.M. & Gouardères, G. (2010). The Affective Tutoring System. Expert Systems with Applications, 37(4), 3013-3023. Elsevier Ltd.

Bennett, J. (2003). Evaluation Methods in Research. London: Continuum.

Blessing, S.B., Gilbert, S.B., Ourada, S. & Ritter, S. (2009). Authoring Model-Tracing Cognitive Tutors. International Journal of Artificial Intelligence in Education, 19(2), 189–210. IOS Press.

Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4–16. JSTOR.

Bower, M. (2009). Using LAMS to Facilitate a “Pedagogy-first”Approach to Teaching Learning Design. Teaching English with Technology – Special Issue on LAMS and Learning Design, 2(3), 42-52.

Boyd, G.M. (2004). Conversation theory. In D. D. Jonassen (Ed.), Handbook of research on educational communications and technology (2nd ed., pp. 179-198). Mahwah, NJ: Lawrence Erlbaum Associates.

Boyer, K.E., Phillips, R., Wallis,M., Vouk, M. & Lester, J. (2008). Balancing cognitive and motivational scaffolding in tutorial dialogues. In:Woolf, B.P.,Aïmeur, E.,Nkambou, R.,Lajoi, S. (eds.) Proceedings Intelligent Tutoring Systems, pp. 239–249. Springer-Verlag, Berlin.

Boyer, K., Phillips, R., Ingram, A., Ha, E., Wallis, M., Vouk, M. & Lester, J. (2011). Investigating the Relationship Between Dialogue Structure and Tutoring Effectiveness: A Hidden Markov Modeling Approach. International Journal of Artificial Intelligence in Education, 21, 65-81.

Bransford, J.D., Brown, A.L. & Cocking, R.R. (Eds.). (2004). How People Learn: Brain, Mind, Experience, and School (pp. 1-374). Washington, D.C.: National Academy Press.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77–101. Taylor & Francis.

Briggs, S.R. & Cheek, J.M. (1986). The Role of Factor Analysis in the Development and Evaluation of Personality Scales. Journal of Personality, 54, p.106-48.

Brown, J.S., & Burton, R.R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive science, 2, 155-192.

Brown, J.S., Burton, R.R. & DeKleer, J. (1982). Pedagogical , natural language and knowledge engineering techniques in SOPHIE I , II and Ill. In D. H. Sleeman & J. S. Brown (Eds.), In Intelligent Tutoring Systems (pp. 272-282). London: Academic Press.

Brown, J.S., Collins, A. & Duguid, P. (1989). Situated Cognition and the Culture of Learning. Educational Researcher, 18(1), 32-42.

Brusilovsky, P.L. (1994). The construction and application of student models in intelligent tutoring systems. Journal of computer and systems sciences international, 32(1), 70–89.

Brusilovsky, P. (1999). Adaptive and intelligent technologies for web-based education. In C. Rollinger and C. Peylo (eds.), Künstliche Intelligenz (4), Special Issue on Intelligent Systems and Teleteaching, 19-25., 2, 1.

Brusilovsky, P. (2003). Developing adaptive educational hypermedia systems: From design models to authoring tools. In T. Murray, S. Blessing, & S. Ainsworth (Eds.), Authoring tools for advanced technology Learning Environments (pp. 377-409). London: Kluwer Academic Publishers.

Brusilovsky, P., Farzan, R. & Ahn, J. (2006). Layered Evaluation of Adaptive Search. In: Workshop on Evaluating Exploratory Search Systems at SIGIR06, August 6-11, 2006 (pp. 11-13). Seattle, WA: USA: ACM Press.

Brusilovsky, P., Karagiannidis, C. & Sampson, D. (2001). The Benefits of Layered Evaluation of Adaptive Applications and Services. In: 1st Workshop on Empirical Evaluation of Adaptive Systems at UM2001 (pp. 1-8). Sonthofen, Germany.

Brusilovsky, P., Sosnovsky, S., Yudelson, M.V., Lee, D.H., Zadorozhny, V. & Zhou, X. (2010). Learning SQL Programming with Interactive Tools: From Integration to Personalization. ACM Transactions on Computing Education (TOCE), 9(4), 1–15. ACM.

Buckler, B. (1996). A learning process model to achieve continuous improvement and innovation. The Learning Organization, 3(3), 31-39.

Bull, S. & Kay, J. (2008). Metacognition and Open Learners Models. The 3rd Workshop on Metacognition and Self-Regulated Learning in Educational Technologies, Montreal, Canada, June 23, 2008.

Burnard, P., Gill, P., Stewart, K., Treasure, E. & Chadwick, B. (2008). Analysing and presenting qualitative data. British dental journal, 204(8), 429-32.

Butz, C.J., Hua, S., & Maguire, R.B. (2004). A Web-Based Intelligent Tutoring System for Computer Programming. IEEE/WIC/ACM International Conference on Web Intelligence (WI'04) (pp. 159-165). IEEE.

Cade, W.L., Copeland, J.L., Person, N.K., & D'Mello, S. (2008). Dialogue modes in expert tutoring. In:Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) Proceedings on Intelligent Tutoring Systems, pp. 470–479. Springer-Verlag, New York, NY.

Cameron, L. (2009). Using LAMS Chat and Forum to Promote Quality Conversations. Teaching English with Technology - Special Issue on LAMS and Learning Design, 2(3), 18-26.

Carberry, S. (2001). Techniques for plan recognition. User Modeling and User-Adapted Interaction, 11, 31-48.

Carbonell, J.R. (1970). AI in CAI: An Artificial-Intelligence Approach to Computer-Assisted Instruction. IEEE Transactions on Man-Machine Systems, 11(4), 190-202.

Carr, B. & Goldstein, I.P. (1977). Overlays: A theory of modelling for computer aided instruction. Massachusetts Institute of Technology. Retrieved from http://18.7.29.232/handle/1721.1/5772

Cassell, C., Buehring, A., Symon, G., Johnson, P. & Bishop, V. (2006). Qualitative management research: a thematic analysis of interviews with stakeholders in the field. ESRC Benchmarking, (June). Retrieved from http://e-space.openrepository.com/e-space/handle/2173/5015

Chalfoun, P. & Frasson, C. (2011). Subliminal Cues While Teaching: HCI Technique for Enhanced Learning. Advances in Human-Computer Interaction, 2011, 1-15.

Chan, Y.H. (2003). Biostatistics 102 : Quantitative Data – Parametric. *Singapore Med. Journal*, *44*(8), 391-396.

Chan, T.W., Roschelle, J., His, S., Kinshuk, K., Sharples, M., Brown, T. & Patton, C. (2006). One-to-one technology-enhanced learning: An opportunity for global research collaboration. Research and Practice in Technology Enhanced Learning, 1(1), 3–29.

Chaouachi, M., Chalfoun, P., Jraidi, I. & Frasson, C. (2010). Affect and mental engagement: towards adaptability for intelligent systems. Proceedings of the 23rd International FLAIRS Conference. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.161.7222&rep=rep1&type=pdf

Cheema, S. & LaViola, J.J. (2012). PhysicsBook: A Sketch-based Interface for Animating Physics Diagrams. Human Factors (pp. 51-60).

Chenail, R.J. (2012). Conducting Qualitative Data Analysis : Qualitative Data Analysis as a Metaphoric Process. The Qualitive Report, 17(1), 248-253.

Chi, M.T.H., Feltovich, P.J. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121–152

Chi, M. & VanLehn, K. (2010). Meta-Cognitive Strategy Instruction in Intelligent Tutoring Systems: How, When, and Why. Educational Technology & Society, 13(1), 25-39.

Chi, M., VanLehn, K., Litman, D. & Jordan, P. (2011). An Evaluation of Pedagogical Tutorial Tactics for a Natural Language Tutoring System: A Reinforcement Learning Approach. International Journal of Artificial Intelligence in Education, 21, 83-113.

Chieu, V., Luengo, V., Vadcard, L. & Tonetti, J. (2010). Student Modeling in Orthopedic Surgery Training: Exploiting Symbiosis between Temporal Bayesian Networks and Fine-grained Didactic Analysis. International Journal of Artificial Intelligence in Education, 20(3), 269–301. IOS Press.

Chin, J.P., Diehl, V.A. & Norman, K.L. (1988). Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. Proceedings of SIGCHI ’88 (pp. 213-218). New York: ACM/SIGCHI.

Chin, W.W., Junglas, I. & Roldán, J. L. (2012). Some considerations for articles introducing new and/or novel quantitative methods to IS researchers. European Journal of Information Systems, 21(1), 1-5.

Ciloglugil, B. & Inceoglu, M. (2010). Exploring the state of the art in adaptive distributed learning environments. In D. Taniar (Ed.), Computational Science and Its Applications–ICCSA 2010 (Vol. Part II, pp. 556–569). Berlin Heidelberg:

Springer-Verlag. Retrieved from http://www.springerlink.com/index/P20N71157R023017.pdf

Cohen, J. (1988). Statistical Power Analysis for the Behavioural Sciences (2nd ed.). USA: Lawrence Erlbaum Associates.

Cohen, J. (1992). A Power Primer. Psychological bulletin, 112(1), 155-159.

Collins, A. (1991a). Cognitive apprenticeship and instructional technology. In L. Idol & B. F. Jones (Eds.), Dimensions of Thinking and Cognitive Instruction (pp. 121–138). Hillsdale, N.J.: Eribaum.

Collins, A. (1991b). The role of computer technology in restructuring schools. Phi Delta Kappan, 73(1), 28–36.

Collins, A., Brown, J.S. & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. American Educator, 15(3), 6–11.

Collins, A., Brown, J.S. & Newman, S. E. (1989). Cognitive Apprenticeship: Teaching the Crafts of Reading, Writing, and Mathematics. In L. B. Resnick (Ed.), Knowing, Learning, and Instruction: Essays in honor of Robert Glaser. (pp. 453-494). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Collins, K.M.T., Onwuegbuzie, A.J. & Jiao, Q.G. (2007). A Mixed Methods Investigation of Mixed Methods Sampling Designs in Social and Health Science Research. Journal of Mixed Methods Research, 1(3), 267-294.

Conati, C. (2009). Intelligent tutoring systems: new challenges and directions. Proceedings of the 14th International Conference on (pp. 2-7). Retrieved from http://www.aaai.org/ocs/index.php/IJCAI/IJCAI-09/paper/viewFile/671/576

Conati, C., Gertner, A.S., Vanlehn, K. & Druzdzel, M.J. (1997). On-Line Student Modeling for Coached Problem Solving Using Bayesian Networks. Proceedings of UM–97, Sixth International Conference on User Modeling (pp. 231–242). Chia Laguna, Italy.

Conati, C., Gertner, A. & Vanlehn, K. (2002). Using Bayesian networks to manage uncertainty in student modeling. User Modeling and User-Adapted Interaction, 12(4), 371–417. Springer.

Conati, C. & VanLehn, K. (1996). Probabilistic plan recognition for cognitive apprenticeship. Proceedings of the eighteenth annual conference of the Cognitive Science Society: July 12-15, 1996, University of California (pp. 403-408). SanDiego, CA. U.S.A.: Lawrence Erlbaum. Retrieved from http://books.google.com/books?hl=en&lr=&id=0S3JeDO8I6sC&oi=fnd&pg=PA403&dq=Probabilistic+Plan+Recognition+for+Cognitive+Apprenticeship&ots=5QBDxzQkrh&sig=QaB3LRd8Bc6Gu87xNXgROLEg-S0

Conati, C. & Vanlehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. International Journal of Artificial Intelligence in Education, 11(4), 389–415.

Conboy, K., Fitzgerald, G. & Mathiassen, L. (2012). Qualitative methods research in information systems: motivations, themes, and contributions. European Journal of Information Systems, 21(2), 113-118. Nature Publishing Group.

Corbett, A. (2001). Cognitive computer tutors: Solving the two-sigma problem. In M. Bauer, P. Gmytrasiewicz, & J. Vassileva (Eds.), User Modeling 2001: Proceedings of the Eighth International Conference (pp. 137-147). New York: Springer. Retrieved from http://www.springerlink.com/index/H6WVCHFGG0U9PDJL.pdf

Corbett, A.T. & Anderson, J.R. (1989). Feedback timing and student control in the LISP Intelligent Tutoring System. In D. Bierman, J. Breuker, & J. Sandberg (Eds.), Artificial Intelligence and Education: Proceedings of the 4th International Conference on AI and Education, 24-26 May, 1989 (pp. 64–72). Amsterdam, Netherlands: IOS. Retrieved from http://act-r.psy.cmu.edu/papers/168/corbett_and_anderson_feedback.pdf

Corbett, A.T. & Anderson, J.R. (2001). Locus of feedback control in computer-based tutoring: impact on learning rate, achievement and attitudes. Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 245–252). ACM. Retrieved from http://portal.acm.org/citation.cfm?id=365111

Corbett, A.T., Koedinger, K.R. & Anderson, J.R. (1997). Intelligent tutoring systems. Handbook of Human-Computer Interaction, Second, Completely Revised Edition in M. Helander, T.K. Landauer, P. Prabhu (Eds.) (2nd ed., Vol. 228, pp. 849-874). Elsevier Science B. V.

Coştu, S., Aydin, S. & Filiz, M. (2009). Students ' conceptions about browser-game-based learning in mathematics education : TTNetvitamin case. Procedia-Social and Behavioral Sciences, 1(1), 1848–1852. Elsevier.

Creswell, J.W. (1998). Qualitative inquiry and research design: Choosing among five traditions. Thousand Oaks, CA: Sage.

Creswell, J.W. & Plano Clark, V.L. (2007). Designing and Conducting Mixed Methods Research. London: Sage Publications, Inc.

Cruzes, D.S. & Dyba, T. (2011). Recommended Steps for Thematic Synthesis in Software Engineering. 2011 International Symposium on Empirical Software Engineering and Measurement, (7491), 275-284. IEEE.

Darlaston-Jones, D. (2007). Making Connections: The Relationship Between Epistemology and Research Methods. The Australian Community Psychologist, 19(1), 19-27.

Dede, C. (2008). Theoretical Perspectives Influencing The Use Of Information Technology. In J. Voogt & G. Knezek (Eds.), International Handbook of

Information Technology in Primary and Secondary Education (Vol. 20, pp. 43-62). Springer Science+Business Media, LLC.

Dennen, V.P. (2004). Cognitive apprenticeship in educational practice: Research on scaffolding, modeling, mentoring, and coaching as instructional strategies. In D. H. Jonassen (Ed.), Handbook of research on educational communications and technology (2nd ed., pp. 813–828). Lawrence Erlbaum Associates Mahwah, NJ.

Dennen, V.P. & Burner, K.J. (2008). The Cognitive Apprenticeship Model in Educational Practice. In J. M. Spector, M. D. Merrill, J. van Merrienboer, & M. P. Driscoll (Eds.), Handbook of research on educational communications and technology (3rd ed., pp. 425-439). New York: Lawrence Erlbaum Associates.

Dennis, C. (2009). Lams As A Cognitive Tool For Teacher Education Students' reflective Thinking. Teaching English With Technology – Special Issue on LAMS and Learning Design, 2(3), 27-41.

Denscombe, M. (2007). The Good Research Guide for small-scale social research projects. 3rd Edition. Open University Press, McGraw-Hill Companies.

Denscombe, M. (2010). The Good Research Guide for small-scale social research projects. 4th Edition. Open University Press, McGraw-Hill Companies.

Dessus, P., Mandin, S. & Zampa, V. (2008). What is Teaching? Cognitive-Based Tutoring Principles for the Design of a Learning Environment. In S. Tazi & K. Zreik (Eds.), International Conference in Human System Learning, Common innovation in e-learning, machine learning and humanoid (ICHSL.6) (pp. 49-55). Toulouse, Paris : France (Europia/IEEE). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.158.515

Dewan, P. (2012). How a language-based GUI generator can influence the teaching of object-oriented programming. Proceedings of the 43rd ACM technical symposium on Computer Science Education (pp. 69–74). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2157159

Di Eugenio, B., Fossati, D., Ohlsson, S. & Cosejo, D. (2009). Towards explaining effective tutorial dialogues. In: Taatgen, N.A., van Rijn, H. (eds.) Proceedings of 31th Annual Conference of the Cognitive Science Society, pp. 1430–1435. Cognitive Science Society, Austin, TX.

Dickey, M. (2008). Integrating cognitive apprenticeship methods in a Web-based educational technology course for P-12 teacher education. Computers & Education, 51(2), 506-518.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), In Three worlds of CSCL. Can we support CSCL? (pp. 61-91.). Heerlen: Open University Nederland. Retrieved from http://hal.archives-ouvertes.fr/hal-00190230/

Direne, A., Bona, L., Sunye, M. S., Castilho, M., Silva, F., Garcia, L., & Scott, D. (2009). Authoring Adaptive Tutoring Systems for Complex Visual Skills. *2009*

Ninth IEEE International Conference on Advanced Learning Technologies, 160-164. IEEE.

Diziol, D., Walker, E., Rummel, N. & Koedinger, K.R. (2010). Using Intelligent Tutor Technology to Implement Adaptive Support for Student Collaboration. Educational Psychology Review, 22(1), 89-102.

Draper, S.W., Brown, M.I., Henderson, F.P. & McAteer, E. (1996). Integrative evaluation: an emerging role for classroom studies of CAL. Computers and Education, 26, (1-3) (1996) 17-32.

du Boulay, B., Avramides, K., Luckin, R., Martinez-Mirón, E., Méndez, G.R. & Carr, A. (2010). Towards Systems That Care: A conceptual framework based on motivation, metacognition and affect. International Journal of Artificial Intelligence in Education, 20(3), 197–229. IOS Press.

du Boulay, B. & Luckin, R. (2001). Modelling human teaching tactics and strategies for tutoring systems. International Journal of Artificial Intelligence in Education, 12, 235-256.

Eagle, M.J. & Barnes, T. (2012). A learning objective focused methodology for the design and evaluation of game-based tutors. Proceedings of the 43rd ACM technical (pp. 99-104). Retrieved from http://dl.acm.org/citation.cfm?id=2157170

Edmondson, R. (2007). Investigating the effectiveness of a telepresence-enabled cognitive apprenticeship model of teacher professional development. Proceedings of 10th international workshop on presence (pp. 265–276). Retrieved from http://therenow.squarespace.com/downloads/Edmondson.pdf

El-Sheikh, E. & Sticklen, J. (1998). A Framework for Developing Intelligent Tutoring Systems Incorporating Reusability. IEA-98-AIE: 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (Vol. 1415). Benicassim, Castellon, Spain: Springer-Verlag.

Enkenberg, J. (2001). Instructional design and emerging teaching models in higher education. Computers in Human Behavior, 17(5-6), 495-506.

Everitt, B. (2006). Medical Statistics from A to Z: A Guide for Clinicians and Medical Students Cambridge University Press: Cambridge, 2 2006.

Fer, S. (2004). Qualitative evaluation of emotional intelligence in-service program for secondary school teachers. The Qualitative Report, 9(4), 562-588.

Fereday, J. & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. International Journal of Qualitative Methods, 5(1), 1–11.

Ferreira, A. & Atkinson, J. (2009). Designing a feedback component of an intelligent tutoring system for foreign language. Knowledge-Based Systems, 22(7), 496-501. Elsevier B.V.

Frankfurt-Nachmias, C. & Nachmias, D. (1996). Research Methods in the Social Sciences. 5th Eds., London: Arnold.

Freedman, R. (2000). What is an intelligent tutoring system. (S. Ali & S. McRoy, Eds.) Intelligence, 11(3), 15–16.

Foster, J.J. (2001). Data Analysis Using SPSS for Windows Versions 8 to 10: A Beginner's Guide (2nd ed.). London: Sage Publications Ltd.

Fowler, F.J. (2002). Survey Research Methods (3rd ed.). London: Sage Publications Ltd.

Fowler, F.J. (2009). Survey Research Methods (4th ed.). London: Sage Publications Ltd.

Gertner, A., Conati, C. & VanLehn, K. 1998. Procedural Help in ANDES: Generating Hints Using a Bayesian Network Student Model. In Proceedings of the Fifteenth National Conference on Artificial Intelligence, 106–111. Menlo Park, Califonia, American Association for Artificial Intelligence.

Gertner, A. & VanLehn, K. (2000). Andes: A coached problem solving environment for physics. In G. Gautheier, C. Frasson, & K. VanLehn (Eds.), Intelligent Tutoring Systems: 5th international Conference, ITS 2000 (pp. 133-142). New York: Springer.

Ghaoui, C. & Janvier, W.A. (2001). Artificial Intelligence in Intelligent Tutoring Systems - An Evaluation, (June), 1-34.

Ghefaili, A. (2003). Cognitive apprenticeship, technology, and the contextualization of learning environments. Journal of Educational Computing, Design & Online Learning, 4(Fall), 1–27.

Giemza, A., Bollen, L., Seydel, P., Overhagen, A. & Hoppe, H.U. (2010). LEMONADE: A Flexible Authoring Tool for Integrated Mobile Learning Scenarios. The 6th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technologies in Education (pp. 73-80). IEEE.

Gilbert, S., Devasani, S., Kodavali, S. & Blessing, S. (2011). Easy Authoring of Intelligent Tutoring Systems for Synthetic Environments. Proceedings of the Twentieth Conference on Behavior Representation in Modeling and Simulation. Retrieved from http://www.shrenikdevasani.com/wp-content/uploads/2011/04/Devasani_2011_BRIMS.pdf

Gillham, B. (2000). Developing A Questionnaire. Contiuum, London.

Gladun, A, Rogushina, J., Garciasanchez, F., Martinezbejar, R. & Fernandezbreis, J. (2009). An application of intelligent techniques and semantic web technologies in e-learning environments. Expert Systems with Applications, 36(2), 1922-1931.

Graesser, A.C., Chipman, P., Haynes, B. C. & Olney, A. (2005). AutoTutor: An Intelligent Tutoring System With Mixed-Initiative Dialogue. IEEE Transactions on Education, 48(4), 612-618.

Graesser, A.C., Conley, M.W. & Olney, A. (2012). Intelligent Tutoring Systems. In K. R. Harris, S. Graham, and T. Urban (Eds.), APA Educational Psychology Handbook, Vol. 3, Chap. 19: Application to Teaching and Learning, pp.451-473.

Graesser, A.C., VanLehn, K., Rose, C.P., Jordan, P.W. & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. AI magazine, 22(4), 39.

Graesser, A.C. & D'mello, S. (2012). Affect-Sensitive Intelligent Tutoring System. US Patent 2012/0052476 A1. Retrieved from http://www.freepatentsonline.com/y2012/0052476.html

Granic, A. (2008). Experience with usability evaluation of e-learning systems. Universal Access in the Information Society, 7(4), 209-221.

Granic, A., Glavinić, V. & Stankov, S. (2004). Usability evaluation methodology for web-based educational systems. 8th ERCIM Workshop" User Interfaces For All." Retrieved from http://crosbi.znanstvenici.hr/datoteka/150759.ERCIM_UI4All_Granic.pdf

Grant, A., Kinnersley, P. & Field, M. (2012). Learning contexts at Two UK medical schools: A comparative study using mixed methods. BMC research notes, 5(1), 153.

Greer, J. & Koehn, G.M. (1995). The peculiarities of plan recognition for intelligent tutoring systems. Proceedings of the IJCAI-95 Workshop on the Next Generation of Plan Recognition Systems. (pp. 54-59). Canada. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.52.3096&rep=rep1&type=pdf

Grundspenkis, J. (2008). Knowledge Creation Supported by Intelligent Knowledge Assessment System. Proceedings of the 12th World Multi-Conference on Systemics, Cybernetics and Informatics, June 29-July 2, 2008 (pp. 135-140). Orlando, Florida, USA.

Grundspenkis, J. & Strautmane, M. (2009). Usage Of Graph Patterns For Knowledge Assessment Based On Concept Maps. Scientific Journal of Riga Technical University, Computer Science, 38(38), 60-71. Versita, Warsaw.

Hammond, L.-D., Austin, K., Orcutt, S. & Rosso, J. (2001). How People Learn: Introduction To Learning Theories. seas.upenn.edu, 1-22. Retrieved from http://www.seas.upenn.edu/~eas285/Readings/Hammond_HowPeopleLearn.pdf

Harper, B. D. & Norman, K. L. (1993). Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5. *Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference* (pp. 224–228). Virginia Beach, VA. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.85.2053&rep=rep1&type=pdf

Hartley, R. (2010). The evolution and redefining of "CAL": a reflection on the interplay of theory and practice. Journal of Computer Assisted Learning, 26(1), 4-17.

Harvey, J. (Eds.) (1998). Evaluation Cookbook. The Learning Technology Dissemination Initiative (LTDI), Edinburgh. Last Accessed: 29th June 2012. Retrieved from: http://www.icbl.hw.ac.uk/ltdi/cookbook/cookbook.pdf

Harwood, T. G. & Garry, T. (2003). An Overview of Content Analysis. The Marketing Review, 3, 479-498.

Hatzilygeroudis, I. & Prentzas, J. (2004). Knowledge representation requirements for intelligent tutoring systems. In J. C. Lester (Ed.), Intelligent Tutoring Systems (pp. 31–44). Berlin Heidelberg: Springer-Verlag.

Hayashi, Y., Bourdeau, J. & Mizoguchi, R. (2009). Using ontological engineering to organize learning/instructional theories and build a theory-aware authoring system. International Journal of Artificial Intelligence in Education, 19(2), 211–252. IOS Press.

Heinze, A. & Heinze, B. (2009). Blended e-learning skeleton of conversation: Improving formative assessment in undergraduate dissertation supervision. British Journal of Educational Technology, 40(2), 294-305.

Heinze, A., Procter, C. & Scott, B. (2007). Use of conversation theory to underpin blended learning. International Journal of Teaching and Case Studies, 1(1/2), 108-120.

Hickey, A.E., & Newton, J.M. (1967). Computer -Assisted Instruction A Survey Of The Literature (2nd Ed.). Newburyport, Mass.: Entelek Inc.

Holland, J. & Childress, M. (2008). Conversation Theory Conceptualized in E-Learning Environments. IGI Global. (pp. 80-89).

Holliday, A. (2007). Doing and Writing Qualitative Research (2nd ed.). London: Sage Publications.

Hou, M., Sobieraj, S., Pronovost, S. & Roberts, S. (2010). Suitable Learning Styles for Intelligent Tutoring Technologies. Defence R&D Canada-Toronto Technical Report, DRDC Toronta TR 2010-073. (pp. 1-74).

Hsieh, H.F. & Shannon, S.E. (2005). Three approaches to qualitative content analysis. Qualitative health research, 15(9), 1277-88.

Hsieh, S. & Hsieh, P. (2001). Intelligent tutoring system authoring tool for manufacturing engineering education. International Journal of Engineering Education, 17(6), 569–579.

Iqbal, A., Oppermann, R., Patel, A. & Kinshuk, K. (1999). A Classification of Evaluation Methods for Intelligent Tutoring Systems. Software Ergonomie'99-Design von Informationswelten.

Isotani, S., Inaba, A., Ikeda, M. & Mizoguchi, R. (2009). An ontology engineering approach to the realization of theory-driven group formation. International Journal of Computer Supported Collaborative Learning, 4(4).

Isotani, S. & Mizoguchi, R. (2008). Theory-Driven Group Formation through Ontologies Theory-driven Group Formation. Intelligent Tutoring Systems (pp. 646-655).

Jackson, G.B. (2002). Intelligent tutoring systems. TechKnowLogia. Knowledge Enterprise, Inc. Retrieved from www.TechKnowLogia.org

Jeremic, Z., Jovanovic, J. & Gasevic, D. (2009). Evaluating an Intelligent Tutoring System for Design Patterns : the DEPTHS Experience. Educational Technology & Society, 12(2), 111-130.

Jeremić, Z., Jovanovic, J. & Gasevic, D. (2012). Student modeling and assessment in intelligent tutoring of software patterns. Expert Systems with Applications, 39(1), 210-222.

Johnson, R.B. & Onwuegbuzie, A.J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. Educational Researcher, 33(7), 14-26.

Johnson, R.B., Onwuegbuzie, A.J. & Turner, L.A. (2007). Toward a Definition of Mixed Methods Research. Journal of Mixed Methods Research, 1(2), 112-133.

Johnson, W.L. (2010). Serious use of a serious game for language learning. International Journal of Artificial Intelligence in Education, 20, 175-195. IOS Press.

Kay, J. (1997). Learner know thyself: student models to give learner control and responsibility. Proceedings of International Conference on Computers. Retrieved from http://www.it.usyd.edu.au/~judy/Homec/Pubs/icce.pdf

Keles, A, Ocak, R., Keles, A. & Gulcu, A. (2009). ZOSMAT: Web-based intelligent tutoring system for teaching–learning process. Expert Systems with Applications, 36(2), 1229-1239.

Keles, A. & Keleş, A. (2010). Intelligent Tutoring Systems : Best Practices. In S. Stankov, V. Glavinic, & M. Rosic (Eds.), Intelligent Tutoring Systems in E-Learning Environments: Design, Implementation and Evaluation (pp. 1-26).

Kerr, A.W., Hall, H.K. & Kozub, S. (2002). Doing Statistics with SPSS. London: Sage Publications Ltd.

Khandaker, N. & Soh, L.K. (2010). Improving group selection and assessment in an asynchronous collaborative writing application. International Journal of Artificial Intelligence in Education, 20(3), 231–268. IOS Press.

Kharya, N., Sharma, S. & Thomas, A. (2010). Fuzzy Grading for Content Management in E-Learning Systems. International Journal of Computer Applications, 1(18), 91-93.

Kim, J., & Gil, Y. (2008). Developing a meta-level problem solver for integrated learners. *Metareasoning: Thinking about Thinking. Papers from the 2008 AAAI Workshop. AAAI Technical Report* (Vol. 8, p. 07). Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Developing+a+Meta-Level+Problem+Solver+for+Integrated+Learners#0

Kimberling, C.L. & Winterstein, A.G. (2008). Validity and reliability of measurement instruments used in research. American Journal of Health-System Pharmacy, Vol. 65, Issue 23, 2276-2284.

Kinshuk (1996). Computer Aided Learning For Entry Level Accountancy Students. Ph.D Thesis, De Montfort University, UK.

Kinshuk, Patel, A. & Russell, D. (2000). A multi-institutional evaluation of Intelligent Tutoring Tools in Numeric Disciplines. Educational Technology & Society, 3(4), 66-74.

Kirkpatrick, D.L. (1979). Techniques for evaluating training programs. Training and Development Journal, 33 (6), 78-92.

Klemm, W. (2002). Software issues for applying conversation theory for effective collaboration via the Internet. Proceedings of the 2002 International Conference. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Software+Issues+for+Applying+Conversation+Theory+For+Effective+Collaboration+Via+the+Internet#0

Kodaganallur, V., Weitz, R.R. & Rosenthal, D. (2006). An Assessment of Constraint-Based Tutors : A Response to Mitrovic and Ohlsson ' s Critique of "A Comparison of Model-Tracing and Constraint-Based Intelligent Tutoring Paradigms." International Journal of Artificial Intelligence in Education., 16(3), 291-321.

Kodaganallur, V., Weitz, R. & Rosenthal, D. (2005). A comparison of model-tracing and constraint-based intelligent tutoring paradigms. International Journal of Artificial Intelligence in Education, 15, 117-144.

Koedinger, K.R., Anderson, J.R., Hadley, W.H. & Mark, M.A. (1997). Intelligent tutoring goes to school in the big city. International Journal of Artificial Intelligence in Education, 8(1), 30–43

Koedinger, K.R. & Corbett, A. (2006). Cognitive Tutors: Technology bringing learning science to the classroom. In Sawyer, K.(eds.), The Cambridge Handbook of the Learning Sciences, Cambridge University Press, pp. 61–78.

Koedinger, K.R, Aleven, V., Heffernan, N., McLaren, B. & Hockenberry, M. (2004). Opening the door to non-programmers: Authoring intelligent tutor behavior by demonstration. In J. C. Lester, R. M. Vicario, & F. Paraguaçu (Eds.), Proceedings of the Seventh International Conference on Intelligent Tutoring Systems (pp. 162-174). Berlin: Springer Verlag. Retrieved from http://www.springerlink.com/index/264mpe819d7brv90.pdf

Koedinger, K.R. & Aleven, V. (2007). Exploring the Assistance Dilemma in Experiments with Cognitive Tutors. Educational Psychology Review, 19(3), 239-264.

Kolmos, A. & Holgaard, J.E. (2008). Learning styles of science and engineering students in problem and project based education. Proceedings of the European Society. for Engineering EducationProc. of Euro. Aalborg, Denmark. Retrieved from http://vbn.aau.dk/files/14952782/learning_styles_of_science.pdf

Koschmann, T. (2001). Revisiting the paradigms of instructional technology. In G. Kennedy, M. Keppell, C. McNaught, & T. Petrovic (Eds.), Meeting at the crossroads. Proceedings of the 18th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education,, Melbourne: The University of Melbourne. (pp. 15–22).

Lacey, A. & Luff, D. (2001). Trent Focus for Research and development in Primary Health Care: An Introduction to Qualitative Data Analysis. Trent Focus, 1-35.

Lane, H.C., & Johnson, W.L. (2008). Intelligent Tutoring and Pedagogical Experience Manipulation in Virtual Learning Environments. In J. Cohn, D. Nicholson, & D. Schmorrow (Eds.), In The PSI Handbook of Virtual Environments for Training and Education. Westport, CT.: Praeger Security International: Westport, CT.

Lane, H. C. & Ogan, A. E. (2009). Virtual Environments for Cultural Learning. 14 th International Conference on Artificial Intelligence in Education, July 6-7th, 2009.

Last, J.M. (ed., 2001). A Dictionary of Epidemiology Oxford University Press, 4 2001.

Lathlean, J. (2006). Qualitative Analysis. In Gerrish, K. and Lacy, A (eds.) The Research Process in Nursing. Pp. 417- 433. Oxford: Blackwell Science.

Laurillard, D. (2002). Rethinking University Teaching – a framework for the effective use of educational technology, London: Routledge.

Lavendelis, E. & Grundspenkis, J. (2009). Design of multi-agent based intelligent tutoring systems. Scientific Journal of Riga Technical University. Computer Sciences, 38(38), 48–59. Versita.

Le, N. T. & Menzel, W. (2008). Towards an Evaluation Methodology of Diagnostic Accuracy for Ill-defined Domains. In Proceedings of the 16th International Conference on Computers in Education. (pp. 117-124). Retrieved from http://www.apsce.net/ICCE2008/papers/ICCE2008-paper101.pdf

Lee, Y., Cho, J. & Choi, B.U. (2006). An Intelligent Tutoring System Based on a Multi-modal Environment for the 300-Certification Program of English Conversation. In M. Ikeda, K. Ashley, & T.-W. Chan (Eds.), Intelligent Tutoring Systems 2006, Lecture Notes in Computer Science (pp. 778–780). Berlin Heidelberg: Springer-verlag.

Leech, N.L., & Onwuegbuzie, A.J. (2009). A typology of mixed methods research designs. Quality & Quantity, 43(2), 265-275.

Lehner, F. & Nosekabel, H. (2002). The role of mobile devices in E-Learning first experiences with a wireless E-Learning environment. Proceedings of the IEEE

International Workshop on Wireless and Mobile Technologies (pp. 2-5). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1039229

Lewis, R. (2001). Cognitive theory, soar. In International Encyclopedia of the Social and Behavioral Sciences. Pergamon, Amsterdam.

Li, D., Zhuying, L. & Bing, Y. (2010). The application of AI technology in Intelligent Tutoring System. Education Technology and Computer (ICETC), 2010 2nd International Conference on (Vol. 1, pp. V1–490). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5529201

Li, H. & Masters, J. (2010). Editorial: E-learning and Knowledge Management in the Early Years: Where Are We and Where Should We Go. Knowledge Management & E-Learning: An International Journal (KM&EL), 1(4), 245.

Liegle, J.O. & Woo, H.G. (2000). Developing Adaptive Intelligent Tutoring Systems : A General Framework and Its Implementations. Proceedings of the ISECON Conference, Nov. 9-12, 2000 (p. 8 p.). Philadelphia.

Lindgaard, G. (1994). Usability Testing and System Evaluation: A guide for designing useful computer systems.UK: Chapman & Hall.

Liu, L., Wang, H., Li, C. & Zhao, C. (2011). Research and application of plan recognition in Intelligent Tutoring System. IT in Medicine and Education (ITME), 2011 International Symposium (pp. 53-56). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6132055

Liu, T. C. (2005). Web-based cognitive apprenticeship model for improving pre-service teachers' performances and attitudes towards instructional planning: Design and field experiment. Educational Technology & Society, 8(2), 136–149.

Longo, D., Katerndahl, D., & Turban, D., et al. (2011). The Research Mentoring Relationship in Family Medicine: Findings From the Grant Generating Project. Family Medicine, 43(4), 240-247.

Lovett, M.C. (2004). Cognitive Theory: ACT. International Encyclopedia of the Social and Behavioral Sciences. pp. 2175-2178.

Lowenthal, P. R. & Leech, N. (2010). Mixed Research and Online Learning: Strategies for Improvement. In T. T. Kidd (Ed.), Online education and adult learning: New frontiers for teaching practices (pp. 1-20). Hershey, PA: IGI Global.

Luckin, R. (2010). Re-designing Learning Contexts: Technology-rich, Learner-Centred Ecologies. London: Routledge, Taylor & Francis Group.

Major, N., Ainsworth, S. & Wood, D. (1997). REDEEM: Exploiting symbiosis between psychology and authoring environments. International Journal of Artificial Intelligence in Education, 8(3/4), 317–340.

Manwaring, G. & Calverley, G. (1998). Directing Your Evaluation. In. J. Harvey (eds.), Evaluation Cookbook. The Learning Technology Dissemination Initiative (LTDI), Edinburgh.

Mara, L., Rodrigues, L. & Carvalho, M. (2004). ITS-I: Intelligent Tutoring System with Integrated Cognition, Emotionand Motivation. Anais do Simpósio Brasileiro de Informática na Educação (Vol. 1, pp. 11–20). Retrieved from http://ceie-sbc.tempsite.ws/pub/index.php/sbie/article/view/302

Marczyk, G., DeMatteo, D., & Festinger, D. (2005). Essentials of Research Design and Methodology, NJ: John Wiley.

Mark, M.A. & Greer, J.E. (1993). Evaluation methodologies for intelligent tutoring systems. Journal of Artificial Intelligence in Education, 4(2-3), 129–153.

Martin, B. (1999). Constraint-based modelling: Representing student knowledge. New Zealand Journal of Computing. Retrieved from http://www.cosc.canterbury.ac.nz/~bim22/paper_nzjc.pdf

Martin, B. & Mitrovic, A. (2000). Tailoring feedback by correcting student answers. In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), Intelligent Tutoring Systems 2000, Lecture Notes in Computer Science, Vol. 1839 (pp. 383-392).

Martin, B.I. (2001). Intelligent Tutoring Systems: The Practical Implementation of Constraint-Based Modelling. Ph.D Thesis, University of Canterbury, New Zealand.

Martin, B. & Mitrovic, A. (2008). Evaluating Intelligent Tutoring Systems with Learning Curves. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.4017&rep=rep1&type=pdf

Martin, B., Mitrovic, A. & Suraweera, P. (2008). ITS Domain Modelling with Ontology. Journal of Universal Computer Science, 14(17), 2758-2776.

Mason, M., 2010. Sample Size and Saturation in PhD Studies Using Qualitative Interviews. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 11(3).

Matsuda, N. & VanLehn, K. (2005). Advanced Geometry Tutor: An intelligent tutor that teaches proof-writing with construction. Proceedings of the 12th International Conference on Artificial Intelligence in Education. (pp. 443- 450.). Amsterdam.

Mayes, T. & de Freitas, S. (2004). Stage 2: Review of e-learning theories, frameworks and models. JISC e-Learning Models Desk Study. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Stage+2:+Review+of+e-learning+theories,+frameworks+and+models#1

McKendree, J. 1990. Effective Feedback Content for Tutoring Complex Skills. Human-Computer Interaction 5:381–413.

McKendree, J., Radlinski, B. & Atwood, M. E. (1992). The GRACE Tutor: A Qualified Success. In Intelligent Tutoring Systems: Second International Conference, eds. C. Frasson, G. Gautheir, and G. I. McCalla, 677–684. Berlin: Springer-Verlag.

McLoughlin, C. (2002). Learning Support In Distance And Networked Learning Environments: Ten Dimensions For Successful Design. Distance Education, 23(2), 149–162.

McMillan, W.J. (2009). Finding a Method to Analyze Qualitative Data: Using a Study of Conceptual Learning. Journal of Dental Education, 73(1), 53–64.

Melis, E. (2005). Choice of Feedback Strategies. *IADIS International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2005).* (pp. 183-189). Retrieved from http://www.iadis.net/dl/final_uploads/200509L022.pdf

Menzel, W. (2006). Constraint-based modeling and ambiguity. International Journal of Artificial Intelligence in, 16(1), 29-63.

Mertler, C.A. & Charles, C.M. (2005). Introduction to Educational Research (5th ed.). USA: Pearson Education, Inc.

Mikic-Fonte, F.A., Burguillo-Rial, J.C., Nistal, M.L. & Fernández-Hermida, D. (2010). A BDI-based intelligent tutoring module for the e-learning platform INES. Frontiers in Education Conference (FIE), 2010 IEEE (p. T1J–1). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5673365

Mikic-Fonte, F.A., Burguillo, J.C. & Nistal, M.L. (2012). An intelligent tutoring module controlled by BDI agents for an e-learning platform. Expert Systems with Applications, 39(8), 7546-7554. Elsevier Ltd.

Miles, M.B. & Huberman, A.M. (1994). Qualitative Data Analysis: An Expanded Sourcebook (2nd ed.). London: Sage Publications Ltd.

Mills, C. & Dalgarno, B. (2007). A conceptual model for game based intelligent tutoring systems. Proceedings ascilite Singapore 2007 (pp. 692-702). Singapore. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.1113&rep=rep1&type=pdf

Miller, S., & Butz, B. (2004). Evaluation of an Interactive Multimedia Intelligent Tutoring System, retrieved March 20, 2009, from http://www.temple.edu/imits/shock/Evaluation2.pdf

Mitrovic, A. (2012). Fifteen years of constraint-based tutors: what we have achieved and where we are going. User Modeling and User-Adapted Interaction, 22(1-2), 39-72.

Mitrovic, A., Martin, B. & Suraweera, P. (2009). ASPIRE: an authoring system and deployment environment for constraint-based tutors. International Journal of Artificial Intelligence in Education, 19, 155-188. IOS Press.

Mitrovic, A. & Weerasinghe, A. (2009). Revisiting ill-definedness and the consequences for ITSs. University of Canterbury. Computer Science and Software Engineering. Retrieved from http://ir.canterbury.ac.nz/handle/10092/2667

Mizoguchi, R., Hayashi, Y. & Bourdeau, J. (2010). Ontology-Based Formal Modeling of the Pedagogical World: Tutors Modeling. In R. Nkambou, J. Bourdeau, and R. Mizoguchi (Eds.), Advances in Intelligent Tutoring Systems. Berlin Heidelberg: Springer-Verlag (pp.229-248).

Moundridou, M. & Virvou, M. (2001a). Authoring and delivering adaptive Web-based textbooks using WEAR. In T. Okamoto, R. Hartley, Kinshuk, & J. P. Klus (Eds.), Proceedings IEEE International Conference on Advanced Learning Technologies; Issues, Achievements, and Challenges—ICALT 2001 (pp. 185-188). Los Alamitos, California: IEEE Comput. Society.

Moundridou, M. & Virvou, M. (2001b). Evaluating the impact of interface agents in an intelligent tutoring systems authoring tool. In N. Avouris & N. Fakotakis (Eds.), Advances in human-computer interaction I: Proceedings of the Panhellenic Conference with International participation in Human-Computer interaction—PC-HCI 2001 (pp. 371–376). Patras, Greece: Typorama Publications. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.6513&rep=rep1&type=pdf

Moundridou, M. & Virvou, M. (2002a). An empirical study for the development of a Web-based authoring tool for Intelligent Tutoring Systems. The 4th International Workshop on Computer Science and Information Technologies (CSIT), September 18-20. Patras, Greece.

Moundridou, M. & Virvou, M. (2002b). Evaluating the instructor support provided by a Web-based authoring tool for building adaptive courses. Proceedings of the IEEE International Conference on Advanced Learning Technologies–ICALT 2002 (pp. 408–413). Retrieved from http://education.aspete.gr/users/mariam/papers/icalt2002.pdf

Moundridou, M. & Virvou, M. (2003a). Analysis and design of a web-based authoring tool generating intelligent tutoring systems. Computers & Education, 40, 157-181.

Moundridou, M. & Virvou, M. (2003b). Usability Evaluation of a Web-based Authoring Tool for Building Intelligent Tutoring Systems. Proceedings of the 10th International. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.6200&rep=rep1&type=pdf

Muldner, K. & Conati, C. (2010). Scaffolding Meta-Cognitive Skills for Effective Analogical Problem Solving via Tailored Example Selection. International Journal of Artificial Intelligence in Education, 20(2), 99–136. IOS Press.

Muller, T.J., Heuvelink, A. & Both, F. (2008). Implementing a Cognitive Model in Soar and ACT-R: A Comparison. Proceedings of Sixth International Workshop" From Agent Theory to Agent Implementation. Retrieved from http://www.ofai.at/research/agents/conf/at2ai6/papers/Muller.pdf

Munoz, K., McKevitt, P., Lunney, T. & Noguez, J. (2006). Virtual Learning Environments and Intelligent Tutoring Systems Survey of current approaches and design methodologies. infm.ulst.ac.uk, 1-27. Retrieved from http://www.infm.ulst.ac.uk/~karlam/PhD/ReviewPaper_Karla.pdf

Munro, A., Johnson, M.C., Pizzini, Q. A., Surmon, D.S., Towne, D.M. & Wogulis, J.L. (1997). Authoring simulation-centered tutors with RIDES. International Journal of Artificial Intelligence in Education, 8(3-4), 284–316.

Murray, T. (1993). Formative qualitative evaluation for "Exploratory" ITS research. Journal of Artificial Intelligence and Education, 4 (2/3), 179-207.

Murray, T. (1997). Expanding the Knowledge Acquisition Bottleneck for Intelligent Tutoring Systems. International Journal of Artificial Intelligence in Education, 8, pp. 222-232.

Murray, T. (1998). Authoring Knowledge-Based Tutors: Tools for Content, Instructional Strategy, Student Model, and Interface Design. Journal of the Learning Sciences, 7(1), 5-64.

Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. International journal of artificial intelligence in education, 10(1), 98–129.

Murray, T. (2003a). An Overview of Intelligent Tutoring System Authoring Tools : Updated Analysis Of The State Of The Art. In T. Murray, S. Blessing, & S. Ainsworth (Eds.), Authoring Tools For Advanced Technology Learning Environments. Kluwer Academic Publishers.

Murray, T. (2003b). EON: Authoring Tools for Content, Instructional Strategy, Student Model, and Interface Design. In Murray, T., Blessing, S., & Ainsworth, S. (Eds.), AuthoringTools for Advanced Technology Learning Environment. Kluwer Academic Publishers.

Murray, T. (2003c). MetaLinks : Authoring and Affordances for Conceptual and Narrative Flow in Adaptive Hyperbooks. International Journal of Artificial Intelligence in Education, 13, 199-233.

Murray, T. (2003d). Principles for Pedagogical-Oriented Knowledge Based Tutor Authoring Systems: Lessons Learned and a Design Meta-Model. In T. Murray, S. Blessing, & S. Ainsworth (Eds.), Authoring Tools for Advanced Technology Learning Environments (pp. 439-466). London: Kluwer Academic Publishers.

Navarro, P. (2001). The limits of social conversation: A sociological approach to Gordon Pask's Conversation Theory. Kybernetes, 30(5), 771-790.

Newell, A. (1994). Unified Theories of Cognition. Harvard University Press.

Newell, A. & Simon, H.A. (1972) Human Problem Solving, Prentice-Hall

Nkambou, R., Bourdeau, J. & Mizoguchi, R. (2010). Introduction: What are Intelligent Tutoring Systems, and Why This Book? In R. Nkambou, J. Bourdeau, and R.

Mizoguchi (Eds.), Advances in Intelligent Tutoring Systems. Berlin Heidelberg: Springer-Verlag (pp.1-14).

Nkambou, R., Bourdeau, J. & Psyche, V. (2010). Building Intelligent Tutoring Systems: An Overview. In R. Nkambou, J. Bourdeau, and R. Mizoguchi (Eds.), Advances in Intelligent Tutoring Systems. Berlin Heidelberg: Springer-Verlag (pp. 361-376).

Nkambou, R., Frasson, C. & Gauthier, G. (2003). CREAM-TOOLs: An Authoring Environment for Knowledge Engineering in Intelligent Tutoring Systems. In T. Murray, S. Ainsworth, & S. Blessing (Eds.), Authoring Tools for Advance Technology Learning Environments (pp. 269-308).

Nwana, H. (1990). Intelligent tutoring systems: an overview. Artificial Intelligence Review, 4(4), 251-277.

Ohlsson, S. (1992). Constraint-based student modeling. Artificial Intelligence in Education 3(4), 429–447.

Ohlsson, S. (1996a). Learning from error and the design of task environments. International Journal of Educational Research, 25(5), 419-448.

Ohlsson, S. (1996b). Learning from performance errors. Psychological Review, 103(2), 241-262.

Ohlsson, S. & Mitrovic, A. (2006). Constraint-based knowledge representation for individualized instruction. Computer Science and Information Systems, 3(1), 1-22.

Ohlsson, S., DiEugenio, B., Chow, B., Fossati, D., Lu, X. & Kershaw,T. (2007). Beyond the code-and-count analysis of tutoring dialogues. In:Luckin, R.,Koedinger, K.,Greer, J. (eds.) Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work, pp. 349–356. IOS Press, Amsterdam.

Oliver, M. & Conole, G. (1998). Selecting a Methodology. In. J. Harvey (eds.), Evaluation Cookbook. The Learning Technology Dissemination Initiative (LTDI), Edinburgh.

Olney, A.M., Graesser, A.C. & Person, N.K. (2012). Question Generation from Concept Maps. Dialogue and Discourse, 3(2), 75-99.

Omoteso, K. (2006). The Impact of Information and Communication Technology on Auditing. Ph.D Thesis, De Montfort University, UK.

Onwuegbuzie, A.J. & Collins, K.M.T. (2007). A typology of mixed methods sampling designs in social science research. The Qualitative Report, 12(2), 281–316.

Onwuegbuzie, A.J. & Leech, N.L. (2007). Sampling designs in qualitative research: Making the sampling process more public. The Qualitative Report, 12(2), 238–254.

Onwuegbuzie, A.J. & Leech, N.L. (2006). Linking research questions to mixed methods data analysis procedures. The Qualitative Report, 11(3), 474-498.

Oppenheim, A.N. (1992). Questionnaire Design, Interviewing and Attitude Measurement, New Edition. London: Continuum.

Padayachee, I. (2002). Intelligent tutoring systems: Architecture and characteristics. In G. Wells & J. McNeill (Eds.), Proceedings of the 32nd Annual SACLA Conference.(Cité p. 2) (pp. 1-8). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.106.5108&rep=rep1&type=pdf

Pallant, J. (2007). SPSS Survival Manual: A Step by Step Guide to Data Analysis using SPSS for Windows. 3rd ed. London: Open University Press.

Pallant, J. (2010). SPSS Survival Manual: A Step by Step Guide to Data Analysis using SPSS for Windows. 4th ed. London: Open University Press.

Paramythis, A., Weibelzahl, S. & Masthoff, J. (2010). Layered evaluation of interactive adaptive systems: framework and formative methods. User Modeling and User-Adapted Interaction, 20(5), 383-453.

Park, H.M. (2008). Univariate analysis and normality test using SAS, STATA, and SPSS. *University Information Technology Services, Center for Statistical and Mathematical Computing, Indiana University*. Bloomington. Retrieved from http://indiana.edu/~statmath/stat/all/normality/normality.pdf

Parscal, T., & Hencmann, M. (2008). Cognitive apprenticeship in online learning. 24th Annual Conference on Distance Teaching & Learning (pp. 1-19). Retrieved from http://www.uwex.edu/disted/conference/Resource_library/proceedings/08_12686.pdf

Pask, G. (1975). Conversation, Cognition and Learning. Amsterdam and New York: Elsevier.

Pask, G. (1976a). Conversational techniques in the study and practice of education. British Journal of Educational Psychology, 46(1), 12–25. Wiley Online Library.

Pask, G. (1976b). Conversation Theory, Applications in Education and Epistemology, Elsevier.

Pask, G. (1976c). Styles and strategies of learning. British Journal of educational Psychology, 46, 128-148.

Pask, G. (1988). Learning strategies, teaching strategies, and conceptual or learning style. In R. D. Schmeck (Ed.), Learning strategies and learning styles (pp. 83–99). New York: Plenum Publishing Corporation.

Pask, G., Scott, B. & Kallikourdis, D. (1973). A theory of conversations and individuals exemplified by the learning process on CASTE. Int. Journal of Man-Machine Studies, Vol. 5, pp. 443-566.

Patel, Cook & Spencer,(2003). Byzantium for learning accounting in English Business Schools. LTSN Business Management Accountancy.

Patel, A. & Kinshuk. (1997). Intelligent Tutoring Tools in a Computer-Integrated Learning Environment for Introductory Numeric Disciplines. Innovations in Education and Teaching International, 34(3), 200-207.

Patel, A., Kinshuk, K. & Russell, D. (2000). Intelligent Tutoring Tools for Cognitive Skill Acquisition in Life Long Learning. Educational Technology & Society, 3(1), 32-40.

Patel, A., Scott, B. & Kinshuk, K. (2001). Intelligent tutoring: from SAKI to Byzantium. Kybernetes, 30(5/6), 807-819.

Paternostro-Sluga, T., Grim-Stieger, M., Posch, M., Schuhfried, O., Vacariu, G., Mittermaier, C., Bittner, C., et al. (2008). Reliability and validity of the Medical Research Council (MRC) scale and a modified scale for testing muscle strength in patients with radial palsy. Journal of rehabilitation medicine : official journal of the UEMS European Board of Physical and Rehabilitation Medicine, 40(8), 665-71.

Peebles, D. & Banks, A.P. (2010). Modelling dynamic decision making with the ACT-R cognitive architecture. Journal of Artificial General Intelligence, 2(2), 52-68.

Penger, S. & Tekavcic, M. (2009). Testing Dunn & Dunn's And Honey & Mumford's Learning Style Theories: The Case Of The Slovenian Higher Education System. Management, 14(2), 1–20.

Peña, A, Sossa, H. & Gutierrez, A. (2007). Cognitive maps: an overview and their application for student modeling. Computación y Sistemas, 10(3), 230-250.

Peña, A. & Sossa, H. (2004). Using Cognitive Maps to Develop a Student Model. Proc. Iberamia 2004, Ibero-american Workshops on Artificial Intelligence. National Institute of Astrophysics, Optics and Electronics. November 22-23th, 2004 (pp. 435-444). Puebla, Mëxico.

Perez-Marin, D. & Pascual-Nieto, I. (2010). Showing Automatically Generated Students' Conceptual Models to Students and Teachers. International Journal of Artificial Intelligence in Education, 20(1), 47–72. IOS Press.

Phobun, P. & Vicheanpanya, J. (2010). Adaptive intelligent tutoring systems for e-learning systems. Procedia - Social and Behavioral Sciences, 2(2), 4064-4069.

Piech, C., Sahami, M., Koller, D., Cooper, S. & Blikstein, P. (2012). Modeling how students learn to program. Proceedings of the 43rd ACM technical symposium on Computer Science Education - SIGCSE '12, 153. New York, New York, USA: ACM Press.

Pong, P. & Challa, S. (2007). Empirical analysis of generalised uncertainty measures with dempster shafer fusion. *10th International Conference on Information Fusion* (pp. 1-9). IEEE.

Pongsuwan, S., Hoksuwan, S., In-udom, W. & Chalakbang, W. (2011). Development of an E-learning Model Based on the Meaningful Learning Process through a Constructivist Theory. European Journal of Social Sciences, 24(4), 625-633.

Popescu, E. (2009). Diagnosing students’ learning style in an educational hypermedia system. Cognitive and Emotional Processes in Web-based Education: Integrating Human Factors and Personalization. Advances in Web-Based Learning Book Series, IGI Global, (13), 187–208.

Pozzebon, E., Cardoso, J., Bittencourt, G., & Hanachi, C. (2007). A group learning management method for intelligent tutoring systems. *INFORMATICA*, *31*, 191-199.

Quinton, S.R. (2010). A Cybernetic Design Methodology for 'Intelligent' Online Learning Support. In J. Yao (Ed.), Web-based Support Systems, Advanced Information and Knowledge Processing (pp. 97-124). London: Springer-Verlag London Limited.

Reiser, B.J., Copen, W.A., Ranney, M., Hamid, A. & Kimberg, D.Y. (2002). Cognitive and Motivational Consequences of Tutoring and Discovery Learning. Cognition and Instruction. Forthcoming.

Riccucci, S. (2008). Knowledge Management in Intelligent Tutoring Systems. Doctoral Thesis, Universit`a di Bologna, Padova.

Rickel, J.W. (1989). Intelligent computer-aided instruction: a survey organized around system components. IEEE Transactions on Systems, Man, and Cybernetics, 19(1), 40-57.

Rindermann, H. (2002). Evaluation. In H. H. Adelsberger, B. Collis, & J. M. Pawlowski (Eds.), Handbook on Information Technologies for Education and Training (pp. 309-329).

Ritter, F.E. (2009). Two Cognitive Modeling Frontiers : Emotions and Usability. Transactions of the Japanese Society for Artificial Intelligence, 24(2), 241-249.

Ritter, S., Anderson, J.R., Koedinger, K.R. & Corbett, A. (2007). Cognitive tutor: applied research in mathematics education. Psychonomic bulletin & review, 14(2), 249-55.

Rodrigues, M., Novais, P. & Santos, M. F. (2005). Future challenges in intelligent tutoring systems-a framework. m-ICTE2005 3rd International Conference on Multimedia and Information and Communication Technologies in Education (pp. 1-5). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.106.7943&rep=rep1&type=pdf

Roll, I., Aleven, V., McLaren, B.M. & Koedinger, K.R. (2007). Designing for metacognition—applying cognitive tutor principles to the tutoring of help seeking. Metacognition and Learning, 2(2-3), 125-140.

Roque, A. & Traum, D. (2007). A model of compliance and emotion for potentially adversarial dialogue agents. Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue (pp. 35-38). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.63.6674&rep=rep1&type=pdf

Ross, S.M. & Morrison, G.R. (2004). Experimental research methods. In D. H. Jonassen (Ed.), Handbook of research for educational communications and technology (2nd ed., pp. 1021-1043). Mahwah, NJ: Lawrence Erlbaum Associates.

Rudman, P., Sharples, M. & Baber, C. (2002). Supporting learning in conversations using personal technologies. In Proceedings of European Workshop on Mobile and Contextual Learning (pp. 44-46). Birmingham, UK, The University of Birmingham, UK. Retrieved from http://www.eee.bham.ac.uk/handler/epapers/mlearn_rudman.pdf

Ruxton, G.D. & Colegrave, N. (2006). Experimental design for the life sciences. 2nd ed. New York: Oxford University Press.

Ryan, R. & Deci, E. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemporary educational psychology, 25(1), 54-67.

Saettler, P. (1990). The Evolution of American Educational Technology. Englewood, Colorado: Libraries Unlimited, Inc., p. 350.

Samuelis, L. (2007). Notes on the Components for Intelligent Tutoring Systems. Acta Polytechnica Hungarica, 4(2), 77–85. Retrieved from http://www.bmf.hu/journal/Samuelis_10.pdf

Sapsford, R. & Jupp, V. (ed.) (2006). Data Collection and Analysis (2nd ed.). London: Sage Publications Ltd.

Scheuer, O., McLaren, B.M., Loll, F. & Pinkwart, N. (2010). Automated Analysis and Feedback Techniques to Support and Teach Argumentation: A Survey. In N. Pinkwart & B. M. McLaren (Eds.), Educational Technologies for Teaching Argumentation Skills. Bentham Science Publishers.

Schiaffino, S., Garcia, P. & Amandi, A. (2008). eTeacher: Providing personalized assistance to e-learning students. Computers & Education, 51(4), 1744-1754.

Schmid, R., DeSimone, C. & McEwen, L. (2001). Supporting the learning process with collaborative concept mapping using computer-based communication tools and processes. Educational Research and Evaluation, 7(2–3), 263–283.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), Perspectives of curriculum evaluation, 39-83. Chicago, IL: Rand McNally.

Scott, B. (2000). CASTE revisited: Principles of course design in a hypertext environment. Information Services and Use, 117-124.

Scott, B. (2001a). Conversation Theory: A Constructivist, Dialogical Apparoach to Educational technology. Cybernetics & Huamn Knowing, 8(4), pp. 25-46.

Scott, B. (2001b). Gordon Pask's conversation theory: A domain independent constructivist model of human knowing. Foundations of Science, 6(4), 343-360.

Scott, B. (2007). "The Cybernetics of Gordon Pask", in Gordon Pask, Philosopher Mechanic: An Introduction to the Cybernetician, R. Glanville and K.H. Muller (eds.), editionechoraum, WISDOM, Vienna, pp.29-52.

Scott, B. (2008). "Woking with Gordon (1967-1978): developing and applying conversation theory", in Handbook of Conversation Design for Instruction, R. Luppicini (ed.), Idea Group, New York, pp. 19-34.

Scott, B., & Cong, C. (2007). Knowledge and task analysis methods for course design. *Readings in Technology in Education: Proceedings …*, *2*(3), 64–79. Retrieved from http://journals.ufv.ca/rr/RR23/article-PDFs/5-scott-cong.pdf

Scott, B., & Cong, C. (2008). "Analysing and representing 'conversational domains'" Proceeding of the 19th European Meeting on Cybernetics and Systems Research, University of Vienna, March 2th-28th, 2008, pp.132-137.

Scott, B., & Cong, C. (2009). Evaluating course design principles for multimedia learning materials. In Proceedings of Defence Academies & Colleges International Conference (pp.127-139). Stellenbosch University, Australia, 25th-27th March, 2009.

Scott, B., & Cong, C. (2010). Evaluating course design principles for multimedia learning materials. *Campus-Wide Information Systems, 27*(5), 280-292.

Self, J. (1990a). Bypassing the intractable problem of student modelling. In C. Frasson & G. Gauthier (Eds.), Intelligent tutoring systems: At the crossroads of artificial intelligence and education (pp. 107–123).

Self, J. (1990b). Theoretical foundations for intelligent tutoring systems. Journal of Artificial Intelligence in Education, 1(4), 3–14.

Self, J. (1994a). Formal approaches to student modelling. In G. I. McCalla & J. Greer (Eds.), Student Modelling: the key to individualized knowledge-based instruction (pp. 295–352). Berlin: Springer-Verlag.

Self, J. (1994b). The role of student models in learning environments. In Transactions of the Institute of Electronics, Information and Communication Engineers (E77-D), 1, 3-8.

Self, J. (1999). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. International Journal of Artificial Intelligence in Education, 10(3-4), 350–364.

Sharples, M. (2002). Disruptive devices: mobile technology for conversational learning. International Journal of Continuing Engineering Education and Lifelong Learning, 12(5/6), 504.

Sharples, M. (2005). Learning as conversation: Transforming education in the mobile age. Proceedings of Conference on Seeing, Understanding, Learning in the Mobile Age (pp. 147–152). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.4715&rep=rep1&type=pdf

Sharples, M., Corlett, D. & Westmancott, O. (2002). The Design and Implementation of a Mobile Learning Resource. Personal and Ubiquitous Computing, 6(3), 220-234.

Sharples, M., Taylor, J. & Vavoula, G. (2010). A Theory of Learning for the Mobile Age. Medienbildung in neuen Kulturräumen. VS Verlag für Sozialwissenschaften (Ed. by Bachmair, B.), pp. 87-99. Accessed on 06/03/2012 at http://dx.doi.org/10.1007/978-3-531-92133-4_6

Shute, V.J. (2008). Focus on Formative Feedback. Review of Educational Research, 78(1), 153-189.

Shute, V.J., & Psotka, J. (1996). Intelligent Tutoring Systems: Past, Present, and Future. In D. Jonassen (Ed.), Handbook of research on educational communications and technology (pp. 570-600.). NY: Macmillan.

Shute, V.J. & Zapata-Rivera, D. (2012). Adaptive Educational Systems. In P. J. Durlach & A. M. Lesgold (Eds.), Adaptive Technologies for Training and Education (pp. 1-35).

Siang, A.C. & Rao, R.K. (2003). Theories of learning: A computer game perspective. Proceedings of the IEEE. IEEE Computer Society. Retrieved from http://www.computer.org/portal/web/csdl/doi/10.1109/MMSE.2003.1254447

Siddappa, M., Manjunath, A. & Kurian, M. (2009). Design, Implementation and Evaluation of Intelligent Tutoring System for Numerical Methods (ITNM). International Conference on Computational Intelligence and Software Engineering, CiSE 2009 (pp. 1–7). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5362751

Siemens, G. & Tittenberger, P. (2009). Handbook of emerging technologies for learning. Image (Rochester, N.Y.). University of Manitoba.

Silverman, D. (2005). Doing Qualitative Research. 2nd ed. London: Sage Publications.

Silverman, D. (2010). Doing Qualitative Research. 3rd ed. London: Sage Publications.

Smith A.S. (1998). Intelligent Tutoring Systems: Personal Notes. Available at http://www.eis.mdx.ac.uk/staffpages/serengul/Tutorial/table.of.contents.htm Accessed: 18/06/2011.

Soloway. E.M. & Johnson, W.L. (1984). Remembrance of blunders past: a retrospective on the development of PROUST. Proceddings of the sixth Cognitive Science Society Conference. Boulder, Colorado.

Span, G. (1993). LITES, an intelligent tutoring system for legal problem solving in the domain of Dutch Civil law. Proceedings of the 4th international conference on Artificial intelligence and law (pp. 76–81). ACM. Retrieved from http://portal.acm.org/citation.cfm?id=158986

Spencer, L., Ritchie, J. & O'Connor, W. (2004). Analysis: Practices, Principles and Processes. In Ritchie J, Lewis J (eds) Qualitative Research Practice. Pp.199-218. London: Sage Publications, 2004.

Strube, G. (2004). Cognitive science: Overview. International Encyclopedia of the Social and Behavioral Sciences, pp. 2158-2166.

Stamper, J., Barnes, T. & Croy, M. (2011). Enhancing the automatic generation of hints with expert seeding. International Journal of Artificial Intelligence in Education, 21, 153-167.

Stankov, S., Glavinic, V. & Grubisic, A. (2004). What is our effect size: Evaluating the educational influence of a web-based intelligent authoring shell. Proc. INES 2004 8th International Conference on Intelligent Engineering Systems, Cluj-Napoca, Faculty of Automation and Computer Science, Technical University of Cluj-Napoca, Romania (pp. 545–550). Retrieved from http://www.pmfst.hr/~ani/radovi/2004INES.pdf

Stathacopoulou, R., Magoulas, G. & Grigoriadou, M. (1999). Neural network-based fuzzy modeling of the student in intelligent tutoring systems. Neural Networks, 1999. IJCNN'99. International Joint Conference on (Vol. 5, pp. 3517–3521). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=836233

Steiner, C. & Hillemann, E. (2010). Refinement and Improvement of Evaluation Framework. *grapple-project.org*. Accessed: 21/10/2011. Retrieved from http://www.grapple-project.org/public-files/deliverables/D8.1b-WP8-Evaluation-Framework-v2.5.pdf

Stoner, G. (2003). Using learning technology resources in teaching management accounting. In: Kaye, R. And Hawkridge, D.G. (eds.), Learning and Teaching for Business: case Studies of Successful Innovation. Kogan Page, London, UK, pp. 138-152.

Suppes, P. (1979). Current Trends in Computer-Assisted Instruction. (M. C. Yovits, Ed.)Advances in Computers, 18, 173-229.

Suraweera, P., Mitrovic, A. & Martin, B. (2010). Widening the Knowledge Acquisition Bottleneck for Constraint-based Tutors. International Journal of Artificial Intelligence in Education, 20(2), 137-173.

Swan, K. (2005). A Constructivist Model For Thinking About Learning Online. (J. Bourne & J. C. Moore, Eds.)Elements of Quality Online Education: Engaging Communities. Needham, MA: Sloan-C.

Sykes, E.R. (2005). Qualitative Evaluation of the Java Intelligent Tutoring System. Journal of Systemics, Cybernetics and Informatics, 3(5), 49-60.

Taatgen, N.A, Lebiere, C. & Anderson, J. (2006). Modeling paradigms in ACT-R. In R. Sun (Ed.), Cognition and multi-agent interaction: From cognitive modeling to social simulation (pp. 29-52). Cambridge, UK: Cambridge University Press.

Taatgen, N. & Anderson, J. (2008). ACT-R. In R. Sun (Ed.), Constraints in Cognitive Architectures (pp. 170-185). Cambridge University Press.

Taatgen, N., & Anderson, J. R. (2010). The Past, Present, and Future of Cognitive Architectures. *Topics in Cognitive Science, 2*(4), 693–704.

Tabachnick, B.G. & Fidell, L.S. (2007). Using Multivariate Statistics (5th ed.). USA: Pearson Education, Inc.

Tashakkori, A. & Teddlie, C. (1998). Mixed Methodology: Combining Qualitative and Quantitative Approaches. London: Sage Publications, Inc.

Tashakkori, A. & Teddlie, C. (2003). Issues and dilemmas in teaching research methods courses in social and behavioural sciences: US perspective. International Journal of Social Research Methodology, 6(1), 61-77.

Taylor, S.J.E. & Siemer, J. (1996). Enhancing simulation education with intelligent tutoring systems. Proceedings of the 28th conference on Winter simulation (pp. 675–680). IEEE Computer Society. Retrieved from http://portal.acm.org/citation.cfm?id=256784

Teddlie, C. & Yu, F. (2007). Mixed Methods Sampling: A Typology With Examples. Journal of Mixed Methods Research, 1(1), 77-100.

Tedman, R. & Tedman, D. (2007). Introduction to the evolution of teaching and learning paradigms. Evolution of Teaching and Learning Paradigms in Intelligent Environment, Studies in Computational Intelligence, 62, 1–6. Springer.

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., Robson, R., et al. (2010). A tutorial on pilot studies: the what, why and how. BMC medical research methodology, 10 (1), p.1-10.

Tinati, R., Halford, S., Carr, L. & Pope, C. (2012). Using Mixed Methods to Track the Growth of the Web: Tracing Open Government Data Initiatives. Analysis. Retrieved from http://eprints.soton.ac.uk/335270/1/WWW12_Web_Science_Track_-_Tracing_Open_Government_Data_Initiatives.pdf

Torrance, H. (2012). Triangulation, Respondent Validation, and Democratic Participation in Mixed Methods Research. Journal of Mixed Methods Research.

Trochim, W. (2001). The Research Methods Knowledge Base. (2nd ed.) USA: Atomic Dog, part of Cengage Learning.

Trochim, W.M.K. & Donnelly, J.P. (2008). The Research Methods Knowledge Base. (3rd ed.) USA: Atomic Dog, part of Cengage Learning.

Turner, P., & Turner, S. (2009). Triangulation in Practice. Virtual Reality, 13(3), 171-181.

Tzouveli, P., Mylonas, P. & Kollias, S. (2008). An intelligent e-learning system based on learner profiling and learning resources adaptation. Computers & Education, 51(1), 224-238.

VanLehn, K. (1987). Learning One Subprocedure per Lesson. Artificial Intelligence, 31(1), 1-40.

VanLehn, K. (1988). Student Modelling. In M. Poison & J. Richardson (Eds.), Foundations of Intelligent Tutoring Systems. (pp. 55-78). Hilisdale, 'NJ: Erlbaum.

VanLehn, K. (2006). The behavior of tutoring systems. International Journal of Artificial Intelligence in Education, 16(3), 227–265. IOS Press.

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist, 46(4), 197-221.

VanLehn, K, Jordan, P., Rose, C. P., Bhembe, D., Bottner, M., Roque, A., Siler, S., et al. (2002). The architecture of Why2-Atlas: A coach for qualitative physics essay writing. In S. A. Cerri, G. Gouarderes, & F. Paraguacu (Eds.), In Proceedings of Intelligent Tutoring Systems Conference (Vol. 2363, pp. 158-167). Springer-Verlag Heidelberg.

VanLehn, K., Jones, R. & Chi, M. (1992). A Model of the Self-Explanation Effect. Journal of the Learning Sciences, 2(1), 1-59.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C.P. (2007). When are tutorial dialogues more effective than reading? Cognitive science, 31(1), 3-62.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., et al. (2005). The Andes physics tutoring system: Lessons learned. International Journal of Artificial Intelligence in Education, 15(3), 147–204. IOS Press.

Vasilyeva, E., Pechenizkiy, M. & Puuronen, S. (2006). The Challenge of Feedback Personalization to Learning Styles in a Web-Based Learning System. Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06), 1143-1144. IEEE.

Vasilyeva, E., Puuronen, S., Pechenizkiy, M. & Rasanen, P. (2007). Feedback adaptation in web-based learning systems. International Journal of Continuing Engineering Education and Life Long Learning, 17(4), 337–357.

Virvou, M. & Moundridou, M. (2000). A web-based authoring tool for algebra-related intelligent tutoring systems. Educational Technology & Society, 3(2), 61–70.

Virvou, M. & Moundridou, M. (2001). Adding an instructor modelling component to the architecture of ITS authoring tools. International Journal of Artificial Intelligence in Education, 12(2), 185–211.

von Glasersfeld, E. (2002). Cybernetics and the Theory of Knowledge. *UNESCO Encyclopedia*. Retrieved from http://www.vonglasersfeld.com/255

Waite M. (2002). Concise Oxford Thesaurus Oxford, England: Oxford University Press, 2 2002.

Wang, F. & Bonk, C.J. (2001). A design framework for electronic cognitive apprenticeship. Journal of Asynchronous Learning Networks, 5(2), 131-151.

Weibelzahl, S. (2003). Evaluation of adaptive systems. PhD Thesis, University of Trier, Germany.

Williams, C., Bower, E.J. & Newton, J.T. (2004). Research in Primary Dental Care, part 6: data analysis. British Dental Journal, 197: 67-73.

Wood, E.J. (2004). Problem-Based Learning : Exploiting Knowledge of how People Learn to Promote Effective Learning. BEE-j, 3(May).

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D. & Picard, R. (2009). Affect-aware tutors : recognising and responding to student affect. Int. Journal on Learning Technology, 4(3/4), 129 - 164.

Woolf, B. & Cunningham, P.A. (1987). Multiple Knowledge Sources in Intelligent Teaching Systems. IEEE Expert, 2(2), 41-54.

Woolf, B.P. (2009). Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning (pp. 1 - 467). San Francisco, CA: Elsevier Inc., Morgan Kauffman.

Yang, A., Kinshuk & Patel, A. (2002). A Plug-Able Web-Based Intelligent Tutoring System. Proceedings of the Xth European Conference on Information Systems (Vol. 83, pp. 1422-1429). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.105.4032&rep=rep1&type=pdf

Young, R.M. (2004). The data learning problem arises also in cognitive architectures other than Soar. Proceedings of the sixth International Conference on Cognitive Modeling (pp. 331–336). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.6423&rep=rep1&type=pdf

Zakharov, K., Mitrovic, A. & Ohlsson, S. (2005). Feedback micro-engineering in EER-tutor. In: Looi, C.-K., McCalla, G., Bredeweg, B., Breuker, J. (eds.) Proceedings of 12th International Conference on Artificial Intelligence in Education, pp. 718–725. IOS Press, Amsterdam.

Zarandi, M.H.F., Khademian, M. & Minaei-Bidgoli, B. (2012). A Fuzzy Expert System Architecture for Intelligent Tutoring Systems: A Cognitive Mapping Approach. Journal of Intelligent Learning Systems and Applications, 04(01), 29-40.

Zelhart, F. & Wallingford, E. (1994). A survey of intelligent tutoring systems and the methods used for effective tutoring. Intelligent Systems Laboratory, 1-35. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.137.1280&rep=rep1&type=pdf

Ziebland, S., & McPherson, A. (2006). Making sense of qualitative data analysis: an introduction with illustrations from DIPEx (personal experiences of health and illness). *Medical Education*, *40*(5), 405-14.

Appendices

Appendix 3.1: A Practical Example of an Author’s Use of ILABS to Construct ILAT for Marginal Costing Topic

The use of ILABS for ITS authoring, for the numerical aspect of applied numerical domains, was demonstrated through the construction of ILATs covering the numerical aspect of marginal costing topic—a management accounting topic. Authors were exposed to the authoring process through the provision of guidelines, pictured in flowchart format as shown in chapter 3—section 3.6.1.1. The referred guidelines and flowchart translated to the stage-by-stage scenario of ILAT authoring, undertaken by one of the authors involved in the research work. The below demonstrated the construction of a marginal costing module (i.e. ILAT), highlighting the steps and some corresponding screenshots.

Sample Authoring Process:

i. Author lunches the ILABS application—double-click ILABS icon on the desktop as shown in fig. 3.15 below (alternatively, one can click the start button on windows, then select ILABS from the options listed under “*All Programs*”).

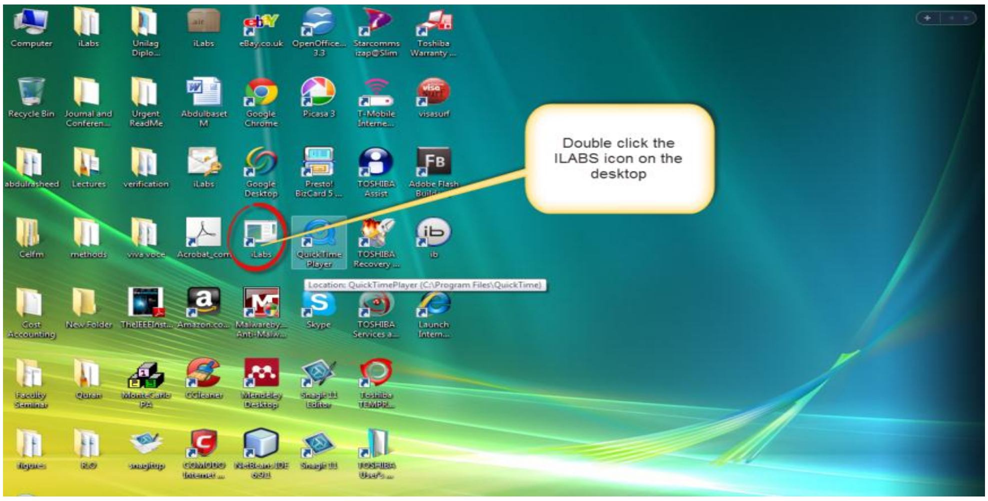


Figure 3.15: Icon for lunching ILABS

ii. From the “File” option—author selects “New” option to display a new module window, then commences construction of a new ILAT (i.e. a new module) as shown in the fig. 3.16 below (or select “Open” to browse and open an existing module/ILAT).

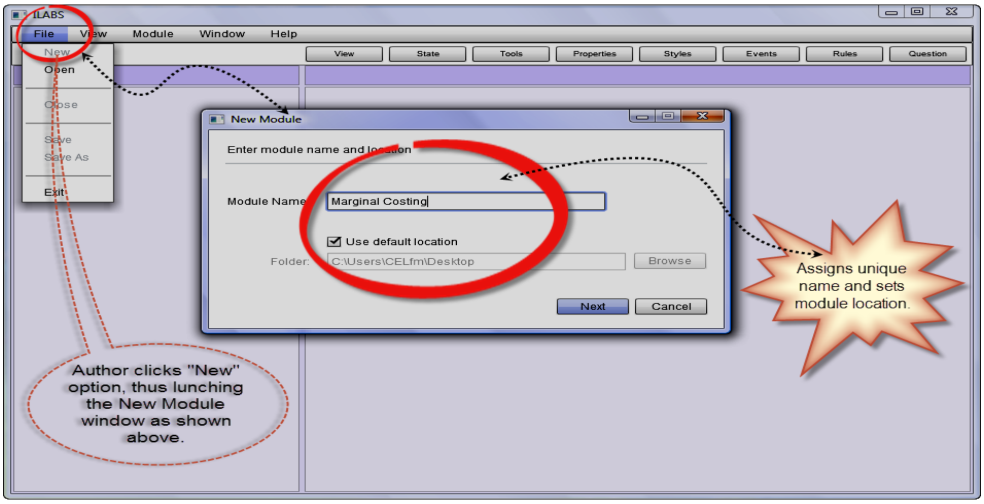


Figure 3.16: File Menu Option for Commencing a New Module

iii. Author assigns a unique name to the new module/ILAT and uses default module location, then clicks "next" button—highlighted—to continue (as shown in fig. 3.16 above)—this display another window where the author selects basic interface and tutoring strategy options (as shown in fig. 3.17 below).

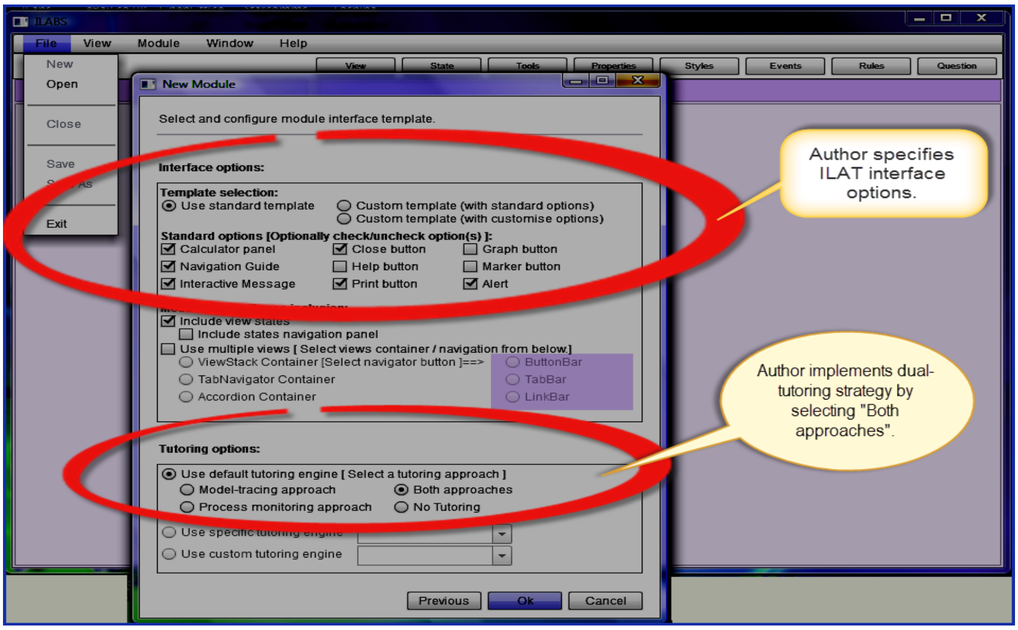


Figure 3.17: Setting Interface and Tutoring Strategy Options

After the above selections, author clicks the highlighted "ok" button to confirm selections. This leads to the display of the ILABS design panels, which includes the tree structure panel where assets dragged from the widgets window are dropped and positioned accordingly and a blank panel meant for future rendering of the tree structure nodes (see fig. 3.18 below).

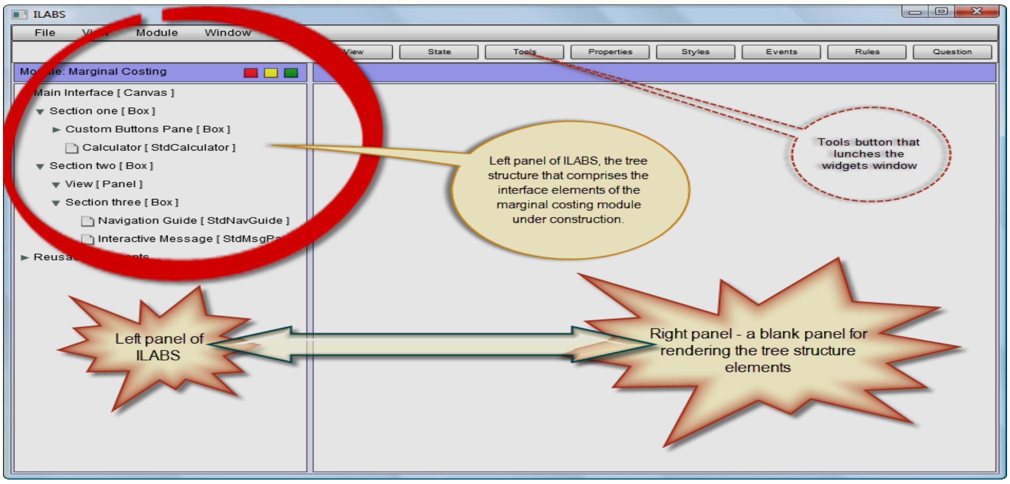


Figure 3.18: The Design Panels of ILABS

iv. Author drag-and-drop assets on to the tree structure on the left hand side of the design interface from the widget window, and position each asset according to how it would appear on the ILAT interface (see fig. 3.18 above).

v. Properties and styles of each asset were set through the "Properties" and "Style" buttons on the ILABS windows respectively (see fig. 3.19 below).

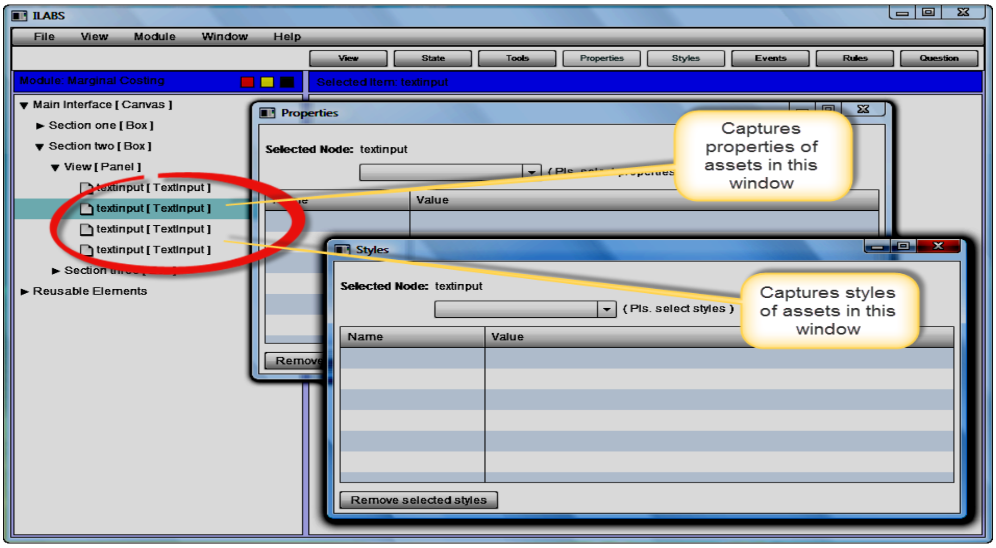


Figure 3.19: Properties and Styles windows

vi. Author clicks "render" under the "Module" option of ILABS window (see fig. 3.20 below). This renders the visual look of the ILAT interface, thus, enhanced judgement in terms of the look and feel of the interface.

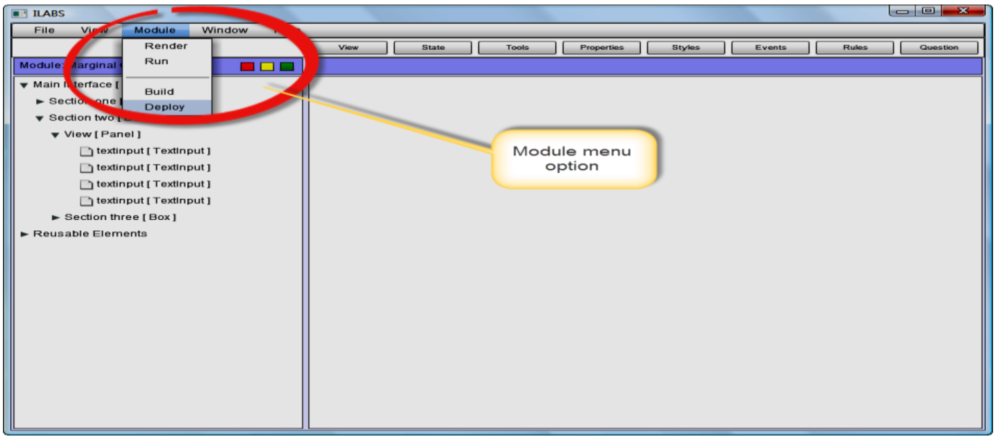


Figure 3.20: The Module menu option on ILABS window

vii. Steps [iv] to [vi] were repeated until author was satisfied with the look and feel of the ILAT interface.

viii. Author clicks the "Rules" button in module menu option on ILABS window (see fig. 3.20 above) to create domain-specific knowledge (i.e. rules that drives the ILAT—this applies to text box assets only—see fig. 3.21 below).

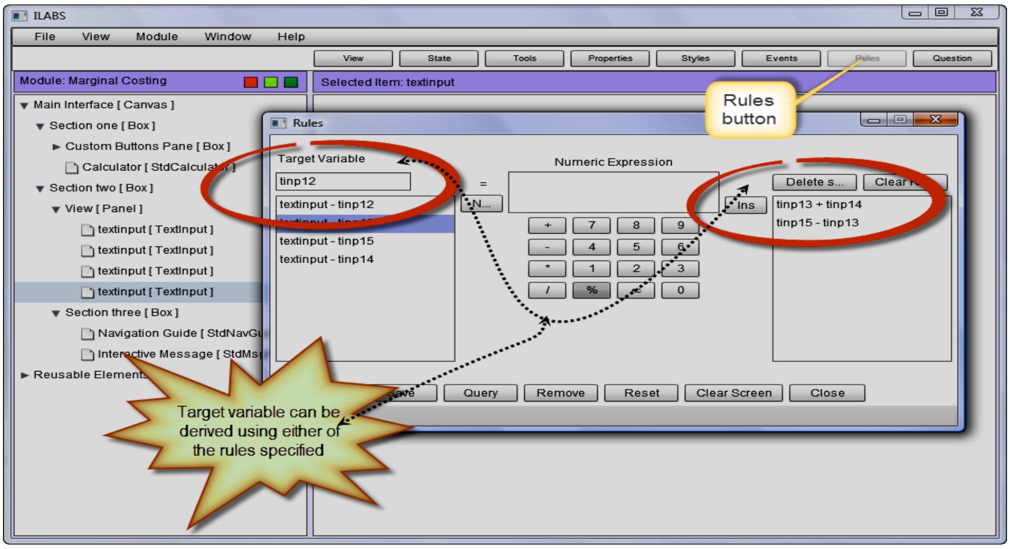


Figure 3.21: Rules Window for Setting Domain-specific Knowledge

ix. Author clicks "Question" button on ILABS' window to display and create problem templates (see fig. 3.22 below). These templates inform the practice problems that were generated during the learning process using the configured ILAT.

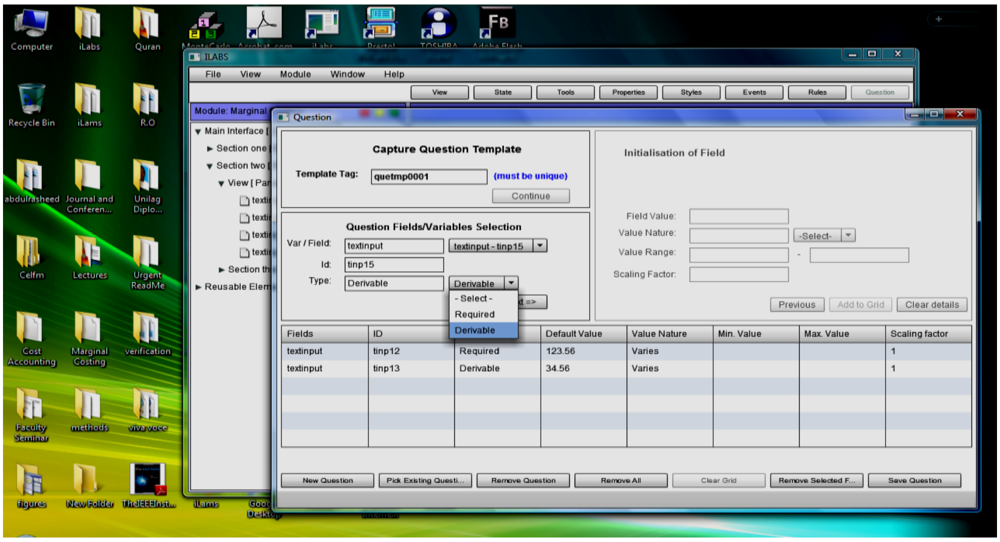


Figure 3.22: Question Template Window

x. Configured ILAT was tested by clicking "run" under the "Module" menu option on ILABS window.

xi. Steps [iv] to [x] were repeated until author was satisfied with configured ILAT workings.

xii. Thereafter, author builds the ILAT and deploy to remote repository using referred options under the "Module" option of the ILABS window.

The above steps gave rise to a configured marginal costing module rendered within the ILABS as depicted in fig. 3.23 below. Each of the authors, involved in this research work, evaluated the authoring process of ILABS as well as the ILAT constructed. Their reaction to the two systems was collected through questionnaires and interviews as indicated in chapter 5 of this thesis.

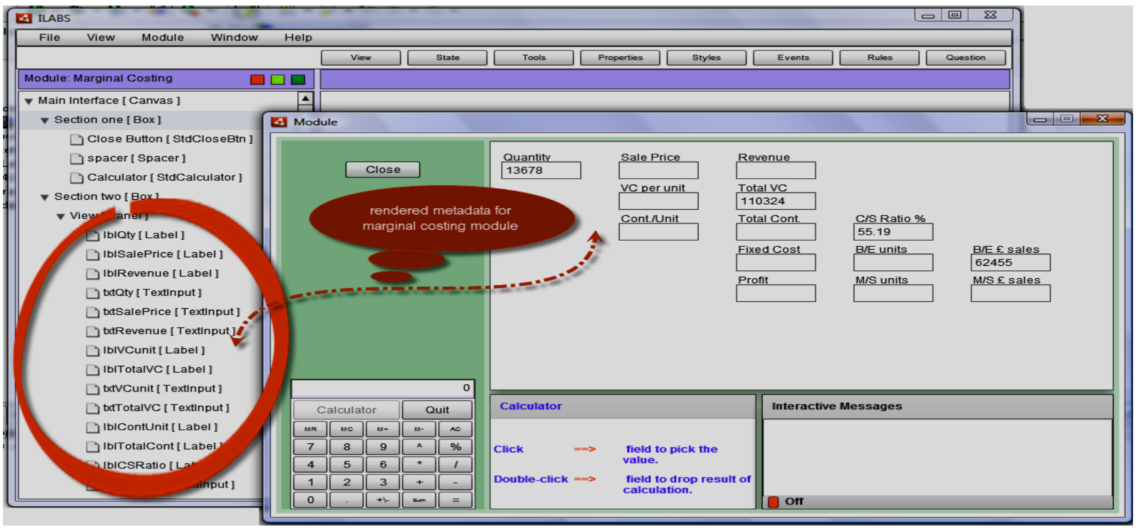


Figure 3.23: Rendered Marginal Costing ILAT

Appendix 3.2: An Exposure of a Learner to Marginal Costing Topic Using ILAT

On the other hand, students utilised/evaluated one of the marginal costing module/ILAT that was constructed by an author. The module implemented dual tutoring strategy but each student was asked to utilise the PM route in order to provide their reaction to its learning impact. The marginal costing ILAT was implemented via ILAMS (see discussion of ILAMS in chapter 3—section 3.6.2). The guidelines below show a typical step-by-step usage of ILAT (via ILAMS) in an exposure/evaluation process (including some screenshots).

Exposure/Evaluation Process:

i. Learner lunches ILAMS application—double-click ILAMS icon on the desktop as shown in fig. 3.24 below (alternatively, one can click the start button on windows, then select ILAMS from the options listed under "*All Programs*").

Figure 3.24: ILAMS Icon

ii. Learner clicks "work offline" to access available modules (i.e. ILATs already downloaded into ILAMS—see fig. 3.25 and 3.26 below).

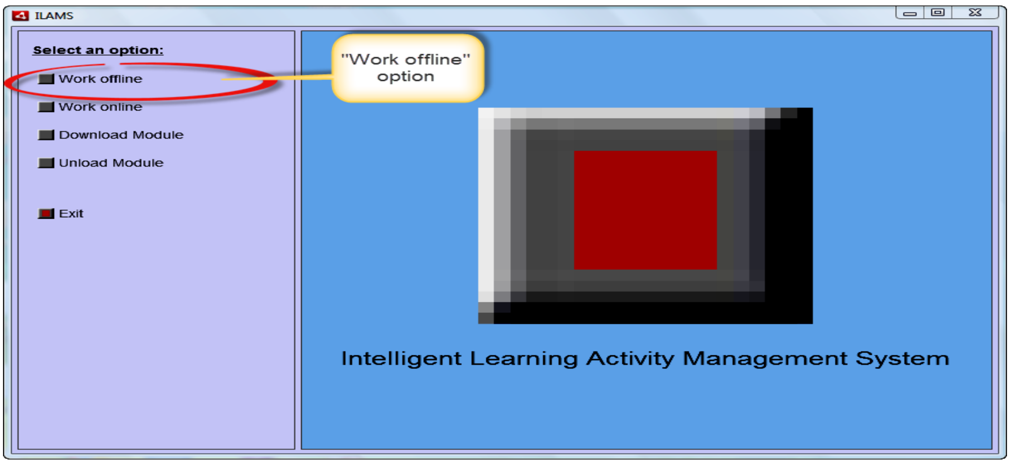


Figure 3.25: Opening Screen for ILAMS

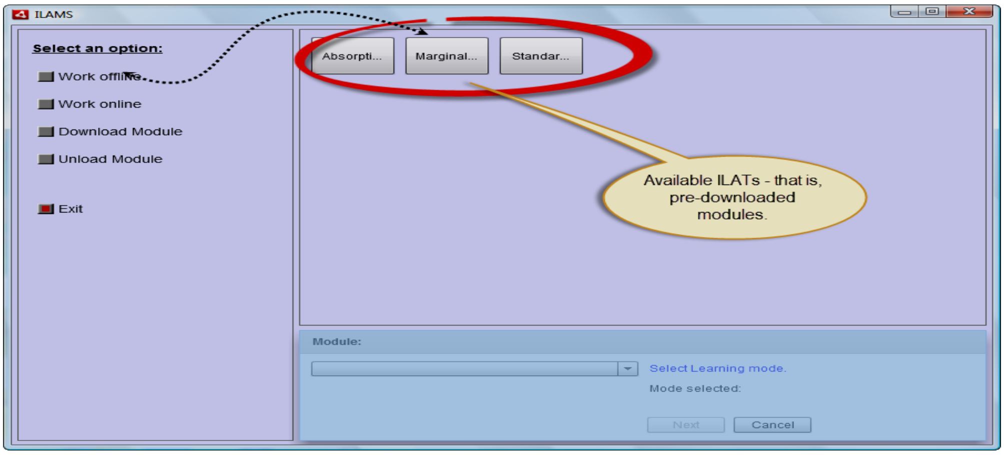


Figure 3.26: ILAMS Showing Downloaded Modules

iii. Learner clicks the “Marginal” costing button; this enables the lower panel, displaying the full name of the module (see fig. 3.27 below). Thereafter, learner selects a learning mode and clicks “Next” button to commence learning (this lunches the marginal costing module rendered in figure 3.23 above).

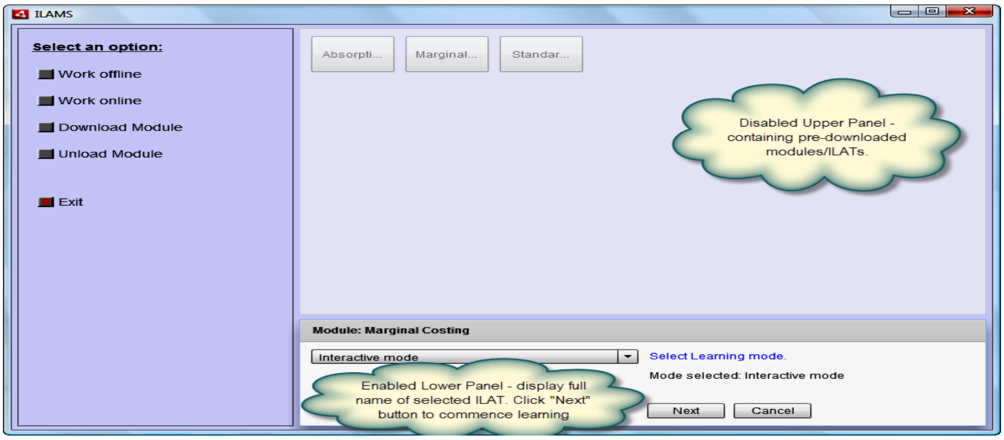


Figure 3.27: Upper and Lower Panel of ILAMS

iv. Click the “on” button of the calculator; this enforces data capturing through the calculator only (however, user can switch the calculator “on/off” during learning, thus alternating between PM route when in “On” state and model-tracing route when in “Off” state).

v. Learner clicks an empty box on the ILAT learning interface to focus the variable to be derived (see fig. 3.23 above). Thereafter, carries out arithmetic operations on the calculator—picking value(s) from box(es) not empty, dropping on calculator and inserting arithmetic operators at appropriate positions, then drops result in focused empty box. This process continues until all the empty boxes are derived with guidance provided by the system. Alternatively, when PM route is not taken, learner carries out arithmetic operations without using the calculator, then enter result directly into empty boxes using the computer keyboard.

The above represents a typical learning process using ILAT via ILAMS during the exposure/evaluation process. Thereafter, the reaction of learners was captured using questionnaires which were analysed as reported in chapter 6 of this thesis.

3

FACULTY OF BUSINESS AND LAW - BUSINESS SCHOOL

ADVANCE APPROVAL OF RESEARCH ACTIVITY INVOLVING HUMAN RESEARCH ETHICS - BUSINESS SCHOOL

Staff/Student Name	Programme/Diet (if relevant)
ADENOWO, ADETOKUNBO A.	PhD

Title of Research Project

THE USE OF ARTIFICIAL INTELLIGENCE IN WEB-BASED INTERACTIVE LEARNING SYSTEM FOR ACCOUNTING STUDENTS.

I agree that in conducting the above research project I will comply with the Market Research Society Code of Conduct as published in their revised July 1999 statement. The rights of respondents are recognised in sections B3 through B8 of the code as per attached to this form. In cases where it is not appropriate to provide written statements, respondents will always receive a verbal statement that their co-operation is voluntary, that anonymity will be preserved, and the purpose for which information is being collected.

I further agree that I will always carry with me and show to respondents my staff/student identification card.

Signature of Researcher/Student — Date: 10/10/07

Signature of Director of Studies/Supervisor — Date: 16/10/08

Signature confirming approval by Designated Officer — Date: 27/10/08

Q:\Postgrad\Graduate Office\General\Forms\RESEARCH ETHICS FORM JULY 2005.doc

QUESTIONNAIRE

Dear respondent,

You have been selected to participate in a survey to voice your opinion on what should characterised a usable Intelligent Learning Activity System Builder (herein referred to as '**<u>Builder</u>**'), that can be used to generate Intelligent Tutoring Systems (i.e. **"<u>eTutors</u>"**) that are meant for learning accounting and finance modules/topics, to compliment traditional classroom teaching.

The exercise is meant for an academic purpose and forms part of the requirements for the degree of Doctor of Philosophy. Therefore, anonymity and confidentiality in completing the questions below is assured.

Please note that you are not in any way compelled to participate; but your participation would be appreciated and considered as consent to be involved in the study.

Thank you

*** GENERAL INSTRUCTION ***

***The questionnaire has three sections, A – D. Please complete all the sections. Kindly, print on blank space(s), and tick one relevant option for items with two or more options, for example*: [√] .**

A. Demographic Characteristics

Your Institution: ....

Your Department: ....

Gender: [ ] Male [ ] Female

Highest Qualification to Date:

[ ] University Graduate [ ] Professional

[ ] Postgraduate [ ] Others (please, specify) ....

Section B: General

	Computer Experience	**Yes**	**No**
1	Can you operate a computer (desktop or laptop)?	[ ]	[ ]
2	Do you know any computer programming language?	[ ]	[ ]
3	Can you write codes in any computer programming langauge?	[ ]	[ ]
4	Have you used any software authoring tool before?	[ ]	[ ]
5	Pls. specify name of the authoring tool you have used (if any):		

Section C: Structured questions

1.1	**Builder system features**	***strongly agree***	***agree***	***neither agree nor disagree***	***disagree***	***strongly disagree***
1	The Builder system gives me option to produce eTutor(s) that enables interactive learning.	[ ]	[ ]	[ ]	[ ]	[ ]
2	The Builder system gives me option to generate eTutor(s) that adapts feedback to students' thinking process.	[ ]	[ ]	[ ]	[ ]	[ ]
3	The builder system allows me to produce eTutor that enable interactive learning as well as adapts feedback to students' thinking process"	[ ]	[ ]	[ ]	[ ]	[ ]

1.2	**Behaviour/Capabilities of generated eTutor(s)**	***strongly agree***	***agree***	***neither agree nor disagree***	***disagree***	***strongly disagree***
1	Response from generated eTutor reflects student's learning process.	[ ]	[ ]	[ ]	[ ]	[ ]
2	Generated eTutor(s) allows conversation (or interaction) between student and system.	[ ]	[ ]	[ ]	[ ]	[ ]
3	Generated eTutor(s) monitors student's problem-solving steps.	[ ]	[ ]	[ ]	[ ]	[ ]

2.1	**Restriction**	***strongly agree***	***agree***	***neither agree nor disagree***	***disagree***	***strongly disagree***
1	The Builder system allows me to generate as many eTutors as I wish.	[ ]	[ ]	[ ]	[ ]	[ ]
2	The Builder system allows me to produce eTutors for different topics (or modules).	[ ]	[ ]	[ ]	[ ]	[ ]
3	The Builder system allows me to generate different variants of an eTutor.	[ ]	[ ]	[ ]	[ ]	[ ]

2.2	**Production Time**	***strongly agree***	***agree***	***neither agree nor disagree***	***disagree***	***strongly disagree***
1	The Builder system allows me to configure and to generate an eTutor within a short span of time.	[ ]	[ ]	[ ]	[ ]	[ ]
2	The Builder system system takes more than 5 hours to configure and generate an eTutor for a 1hour tutorial session.	[ ]	[ ]	[ ]	[ ]	[ ]

2.3	**Special skills**	***strongly agree***	***agree***	***neither agree nor disagree***	***disagree***	***strongly disagree***
1	I need computer programming skill to be able to generate an eTutor from the Builder system.	[ ]	[ ]	[ ]	[ ]	[ ]
2	knowledge of accounting is required to generate meaningful eTutors from the Builder system.	[ ]	[ ]	[ ]	[ ]	[ ]

3.1	**Tutoring strategies**	***strongly agree***	***agree***	***neither agree nor disagree***	***disagree***	***strongly disagree***
1	The Builder system allows me to generate eTutor that monitors student's problem-solving steps.	[ ]	[ ]	[ ]	[ ]	[ ]
2	The Builder system allows me to produce eTutor that monitors student's input value.	[ ]	[ ]	[ ]	[ ]	[ ]
3	The Builder system can produce eTutor(s) that support both features (i.e. monitors problem solving steps and traces student's input value).	[ ]	[ ]	[ ]	[ ]	[ ]

4.1	**Ease of Use and Usability:**	***strongly agree***	***agree***	***neither agree nor disagree***	***disagree***	***strongly disagree***
	Generally, the Builder system can be described as					
1	.......annoying	[ ]	[ ]	[ ]	[ ]	[ ]
2	......confusing	[ ]	[ ]	[ ]	[ ]	[ ]
3	......frustrating	[ ]	[ ]	[ ]	[ ]	[ ]
4	......interesting	[ ]	[ ]	[ ]	[ ]	[ ]
5	......stimulating	[ ]	[ ]	[ ]	[ ]	[ ]
6	......tiresome	[ ]	[ ]	[ ]	[ ]	[ ]
7	......usable	[ ]	[ ]	[ ]	[ ]	[ ]
8	......unpleasant	[ ]	[ ]	[ ]	[ ]	[ ]
9	I feel in control when I am using the system.	[ ]	[ ]	[ ]	[ ]	[ ]

	(Continuation of section - B.4)	***strongly agree***	***agree***	***neither agree nor disagree***	***disagree***	***strongly disagree***
10	Builder system uses terms that is understandable.	[ ]	[ ]	[ ]	[ ]	[ ]
11	Builder system uses terms that is familiar to me.	[ ]	[ ]	[ ]	[ ]	[ ]
12	Builder system needs more introductory explanations.	[ ]	[ ]	[ ]	[ ]	[ ]
13	It is easy to understand the objects on the Builder system's interface.	[ ]	[ ]	[ ]	[ ]	[ ]
14	Builder system is slow	[ ]	[ ]	[ ]	[ ]	[ ]
15	I get what I expect when I click on objects on the Builder system interface.	[ ]	[ ]	[ ]	[ ]	[ ]
16	It is difficult to move around the Builder system.	[ ]	[ ]	[ ]	[ ]	[ ]
17	I feel efficient when using the Builder system.	[ ]	[ ]	[ ]	[ ]	[ ]
18	Builder system can be characterised as innovative.	[ ]	[ ]	[ ]	[ ]	[ ]
19	Overall, I am satisfied with the Builder system.	[ ]	[ ]	[ ]	[ ]	[ ]

Section D: Subjective questions (Please complete the spaces below)

1	The strengths of the Builder system are:	
2	The weaknesses of the Builder system are:	
3	The features/functionalities that could be improved upon are:	
4	The features/functionalities that could be added are:	
5	Other comments, please (e.g. On the Builder system, clarity of questionnaire items etc.)	

Thank you for participating................

DE MONTFORT UNIVERSITY LEICESTER

QUESTIONNAIRE

Dear respondent,

You have been selected to participate in a survey to voice your opinion on using an **Intelligent Learning Activity System** (herein referred to as ‘e**Tutor’**) for learning numerate aspect of accounting & finance modules.

The exercise is meant for an academic purpose and forms part of the requirements for the degree of Doctor of Philosophy. Therefore, anonymity and confidentiality in completing the questions below is assured.

Please note that you are not in any way compelled to participate, but your participation would be highly appreciated and regarded as consent to partake.

Thank you.

*** GENERAL INSTRUCTION ***

The questionnaire has four sections A – D. Please complete all sections.
Please, print on blank space(s), and tick one relevant option for items with two or more options, for example: [√]

A. Demographic Characteristics

Your Institution: ....

Your Department: ....

Gender: [ ] Male [ ] Female

Age: [ ] 16-25 [ ] 26-35 [ ] 36-45 [ ] 46 and above

[ ] Others (Please, specify) ....

Highest Qualification to Date:

[ ] GCE / WASC / NECO / GCSE [ ] Diploma

[ ] A-Levels [ ] Others (pls., specify) ....

Section B: General

	Computer Experience	Yes	No
1	Can you operate a computer (desktop or laptop)?	[ ]	[ ]
2	Do you enjoy using computer at all?	[ ]	[ ]
3	Do you know any computer programming language?	[ ]	[ ]
4	Can you write codes in any computer programming language?	[ ]	[ ]
5	Do you like learning via computer?	[ ]	[ ]
6	Have you used any eTutor (or computer aided learning software) before now?	[ ]	[ ]
7	If question 6 is YES, was it in accounting and/or finance related subject?	[ ]	[ ]
8	Please, write the name of the eTutor you have used before, if any.		

Section C: Structured questions

1.1	Cognitive Process Visible	*strongly agree*	*agree*	*neither agree nor disagree*	*disagree*	*strongly disagree*
1	eTutor accurately capture my thinking process during problem-solving.	[ ]	[ ]	[ ]	[ ]	[ ]
2	eTutor correctly infer my thinking process through my problem-solving steps.	[ ]	[ ]	[ ]	[ ]	[ ]
3	eTutor identified my thinking process through my learning actions.	[ ]	[ ]	[ ]	[ ]	[ ]
4	eTutor's behavour adapts feedback to my thinking process.	[ ]	[ ]	[ ]	[ ]	[ ]
5	eTutor's responses shows it accurately identified my thinking process.	[ ]	[ ]	[ ]	[ ]	[ ]
6	My problem-solving steps were reflected in the eTutor's behaviour.	[ ]	[ ]	[ ]	[ ]	[ ]
7	Responses from the eTutor were relevant to my problem-solving steps.	[ ]	[ ]	[ ]	[ ]	[ ]

1.2	Visibility Through Conversations	*strongly agree*	*agree*	*neither agree nor*	*disagree*	*strongly disagree*

				disagree		
1	My problem-solving steps were identified during interaction with the eTutor.	[ ]	[ ]	[ ]	[ ]	[ ]
2	eTutor provides interface that enables learning through interaction with system.	[ ]	[ ]	[ ]	[ ]	[ ]

2.1	**Timely Feedback**	***strongly agree***	***agree***	***neither agree nor disagree***	***disagree***	***strongly disagree***
1	eTutor provided feedbacks at appropriate time (of need).	[ ]	[ ]	[ ]	[ ]	[ ]
2	eTutor's feedbacks were a little delayed.	[ ]	[ ]	[ ]	[ ]	[ ]

2.2	**Relevant Feedback**	***strongly agree***	***agree***	***neither agree nor disagree***	***disagree***	***strongly disagree***
1	eTutor's feedbacks were relevant to problem-solving task(s).	[ ]	[ ]	[ ]	[ ]	[ ]
2	eTutor's feedback were appropriately framed along task difficulty.	[ ]	[ ]	[ ]	[ ]	[ ]
3	eTutor's feedbacks were supportive/helpful in solving task.	[ ]	[ ]	[ ]	[ ]	[ ]
4	eTutor's feedbacks obstructed learning.	[ ]	[ ]	[ ]	[ ]	[ ]
5	eTutor's feedbacks were appropriately framed along my thinking process.	[ ]	[ ]	[ ]	[ ]	[ ]
6	Feedbacks maps accurately into my thinking process.	[ ]	[ ]	[ ]	[ ]	[ ]
7	eTutor's feedbacks enhanced learning.	[ ]	[ ]	[ ]	[ ]	[ ]

3.1	**Misconceptions:**	***strongly agree***	***agree***	***neither agree nor disagree***	***disagree***	***strongly disagree***
1	eTutor's feedbacks always appropriately address my misunderstanding of task(s).	[ ]	[ ]	[ ]	[ ]	[ ]
2	eTutor's feedbacks shows it accurately identifies my misunderstandings.	[ ]	[ ]	[ ]	[ ]	[ ]
3	eTutor was able to correct my misundertsandings.	[ ]	[ ]	[ ]	[ ]	[ ]
4	Whenever I commit, eTutor did not identify any of my misundertsandings.	[ ]	[ ]	[ ]	[ ]	[ ]

4.1	Learning Effectiveness	*strongly agree*	*agree*	*neither agree nor disagree*	*disagree*	*strongly disagree*
1	The eTutor was useful to my studies.	[ ]	[ ]	[ ]	[ ]	[ ]
2	Learning via the eTutor was beneficial to my studies.	[ ]	[ ]	[ ]	[ ]	[ ]
3	Learning via the eTutor extended classroom teaching.	[ ]	[ ]	[ ]	[ ]	[ ]
4	eTutor was easy to use.	[ ]	[ ]	[ ]	[ ]	[ ]
5	I understood the eTutor's tutoring/logical approach.	[ ]	[ ]	[ ]	[ ]	[ ]
6	I enjoyed learning through the eTutor.	[ ]	[ ]	[ ]	[ ]	[ ]

	Continuation of section C4.1	*strongly agree*	*agree*	*neither agree nor disagree*	*disagree*	*strongly disagree*
7	Learning via the eTutor was confusing.	[ ]	[ ]	[ ]	[ ]	[ ]
8	I learn more through the eTutor than in a regular classroom based tutorial session.	[ ]	[ ]	[ ]	[ ]	[ ]
9	The eTutor is positively challenging.	[ ]	[ ]	[ ]	[ ]	[ ]
10	I learnt better through the eTutor than in a regular classroom based tutorial session.	[ ]	[ ]	[ ]	[ ]	[ ]
11	eTutor stimulates learning.	[ ]	[ ]	[ ]	[ ]	[ ]
12	The eTutor is relevant to my progression or career aspirations.	[ ]	[ ]	[ ]	[ ]	[ ]
13	I would recommend this eTutor to other students.	[ ]	[ ]	[ ]	[ ]	[ ]

5.1	Cognitive Visibility and Learning	*strongly agree*	*agree*	*neither agree nor disagree*	*disagree*	*strongly disagree*
1	Learning will be enhanced if eTutor understands all my problem-solving steps.	[ ]	[ ]	[ ]	[ ]	[ ]
2	Learning will be enhanced if eTutor respond according to my problem-solving steps.	[ ]	[ ]	[ ]	[ ]	[ ]
3	My knowledge of the subject will improve if eTutor can identify my misunderstandings.	[ ]	[ ]	[ ]	[ ]	[ ]

4	My knowledge horizon will broaden, if eTutor's feedbacks are relevant to my misunderstanding.	[ ]	[ ]	[ ]	[ ]	[ ]
5	Learning will improve if my misunderstanding can be detected during interaction with eTutor.	[ ]	[ ]	[ ]	[ ]	[ ]
6	Learning will be enhanced if eTutor provides timely feedbacks on my misunderstandings.	[ ]	[ ]	[ ]	[ ]	[ ]
7	Learning will improve if eTutor does not understands all my problem-solving steps.	[ ]	[ ]	[ ]	[ ]	[ ]
8	My knowledge of the subject will improve if eTutor does not identify my misunderstandings.	[ ]	[ ]	[ ]	[ ]	[ ]
9	Learning will be enhanced if eTutor's feedbacks are not relevant.	[ ]	[ ]	[ ]	[ ]	[ ]

Section D: Subjective questions (Please complete the spaces below)

1	The strengths of the eTutor were:
2	The weaknesses of the eTutor were:

Continuation of Section D (Subjective questions) (Please complete the spaces below)

3	The features/functionalities that could be improved upon are:
4	The features/functionalities that could be added are:
5	Other comments, please (e.g. on eTutor's interface, functionalities, clarity of questionnaire items, etc)

Thank you.

SEMI-STRUCTURED INTERVIEW PROTOCOL
(Intelligent Learning Activity System Builder)

Purpose of the study/interview

To evaluate an Intelligent Learning Activity System Builder (herein referred to as '**Builder**'), which generates online tutoring systems (i.e. **"eTutors"**) that are meant for learning accounting and finance modules/topics, to compliment traditional classroom teaching.

The exercise is meant for an academic purpose and forms part of the requirements for the degree of Doctor of Philosophy. Therefore, anonymity and confidentiality of data collected is assured. Please note that you are not in any way compelled to participate, but participation will be highly appreciated.

<u>Preliminary Questions</u>

1. Please, could you introduce yourself - stating your institution's name, job role, and how long you have been on the role?
2. Have you ever used any online Tutor before now? Could name it/them? How many modules or topics did it cover?

<u>Main Research Questions</u>

Question A:

1. Could you identify some observable features of the Builder system?
2. In what way did the features (from the Builder system) reflect in the eTutor (if at all)?

Question B:

3. What tutoring strategies do (or should) the Builder system provides?
4. With respect to the strategies identified, could you describe the eTutor that could be generated from the Builder system?

Question C:

5. How would you describe the flexibility of the Builder system with respect to the following:
    - the number of eTutors that can be generated?
    - the variants of an eTutor that can be generated?

6. How long would it take to configure and generate an eTutor from the Builder system?
7. What special skill(s) or knowledge would be required in order to be able to use the Builder system?

Question D:

8. How would you describe the Builder system in terms of:
   a. ease of use of its interface(s)?
   b. the usability (or convenience) in accomplishing a task?
   c. Its learning curve in order to be able to use it?
9. In terms of your satisfaction, could you award mark between 0 to 100% for the following:
   (i) ease of use, and
   (ii) usability.

Concluding Questions:

1. Generally, could you identify the strengths and weakness of the Builder system?
2. What should be improved and/or added to make the Builder system to make it usable?
3. What is your view on having eTutors to cover all topics in accounting?
4. Any additional remark?

S/N	Headings	Initial Structure (Before Pilot Study) Number of Items	Final Structure (Used for the Study) Number of Items
1	About You / Preliminary	3	----
2	About Your Institution	9	-----
	Questions:		
3	Preliminary	2	2
4	Proposition 1	2	2
5	Proposition 2	1	2
6	Proposition 3	1	3
7	Proposition 4	8	2
8	Concluding Questions	2	4
Total		***28***	***15***

Appendix 5.1 – Data Cleaning / Preliminary Analysis of Builder Questionnaire Data

A5.1.1 Phases of Data Collection/Analysis

Current appendix augments chapter 5 of this thesis. It explicates the preliminary analysis undertaken in an attempt to provide answers to research question one. To actualise the empirical aspects of research question one, data collected stretched through three phases. Two phases were pilot studies and the third phase was the main (or actual) study. Conduct of two pilot studies, enabled extensive testing of the research instruments to evaluate its reliability and validity. Analysis conducted in this appendix with respect to chapter 5, therefore, covers data collected during the three phases employed.

Two main forms of analysis were undertaken in chapter 5, quantitative and qualitative. Each employed different analytical procedure, as described in the methodology chapter. And it aims to bring into the open, the probabilistic and qualitative meanings, embedded in data gathered via various instruments used in this work. But this appendix addresses the quantitative preliminary analysis only. Within this analytical category, two types of analyses were undertaken: analysis of pilot data sets - to test-run research instruments prior to their deployment in the final data collection phase; analysis of the main study data set – used to address current research propositions.

A5.1.2 Research Instruments / Participants

In accordance with above, the under listed instruments in figure 5.1.1 below were administered to university lecturers in business and allied courses, being the targeted population – the expected users of the builder system (ILABS) developed. The instruments were administered in four higher education providing institutions (code named: Uni. A, B, C and D). The first institution only took part in pilot study I, while the other three institutions took part in both the pilot study II and the main study.

Participants were drawn from key relevant departments in the business schools of the institutions concerned, majorly accounting, banking and finance, and some other departments in the school. Reason was that the context of the research focused on numerical problem-solving in accounting and finance modules. So, lecturers within that disciplines were considered appropriate in this research. Table 5.1.1 below provides the statistics of participants in relation to the three phases of the study. It shows the number of responses received from evaluators with respect to the three data collection phases, institutions involved and instruments used. Note that code naming of institutions, as reflected in the table below, was done for confidentiality purpose.

Instruments:

- Builder Questionnaire (lecturers users only)
- Builder Interview protocol (lecturers users only)

Figure 5.1.1: Research Instruments

Table 5.1.1: Number of Participants in Pilot and Main Studies

Study Type / Institutions		**Questionnaire**	**Interviews**
Pilot 1	Uni.A	4	4
Pilot 2	Uni.B	10	---
	Uni.C	6	---
	Uni.D	8	1
	$\text{Total}_{\text{pilot2}}$	***24***	***1***
Main study	Uni.B	32	5
	Uni.C	25	---
	Uni.D	25	3
$\text{Total}_{\text{main}}$		***82***	***8***

A5.1.3 Data Cleaning / Analysis

This section presents only the quantitative analysis of the study; mainly, the analysis of responses to questionnaire instrument. Pallant (2010) and Mertlier & Charles (2005) provided clue on types of quantitative analyses that can be employed. According to them, prior to extensive analysis that may involve lots of energy, time and other research resources, data should be treated to certain preliminary analysis. It involve checking and cleaning data entered into statistical package; purposely, to eliminate (or at least, reduce drastically) data entering errors. A crucial analytical step, because many statistics are sensitive to slight change due to data capture errors (Pallant, 2007, 2010), allowing it may result in false outcome. It will also identify, where present, missing data. Equally, many of the statistics have certain assumptions that should be met before they can be applied. For example, parametric statistics (e.g. t-test, ANOVA, correlation etc.) require normally distributed data. This then requires, confirmation of the nature of the data before deciding whether to apply such class of statistics or not.

In the light of the above, certain steps were taken prior to the main analysis; those steps yielded the findings discussed in this chapter. This includes data cleaning, factor analysis - where applicable, reliability test and descriptive statistics. The statistics enabled the testing of statistical assumptions, verification of the data entered into IBM SPSS Statistics version 19 and exploration of their nature, prior to further analyses. Through those means, all forms of data capturing errors were identified and eliminated, to prevent negative influences on result used in the core analysis of the research. As well, the steps informed the statistics that were eventually used in the final analysis.

As a first step towards analysis, data entered into IBM SPSS statistics version 19 was screened. This involves checking for data entry errors and outliers; where errors exist, they were removed. This stage of the work was implemented by generating frequency statistics of all variables; its shows frequency of valid cases, missing values, and minimum and maximum values - in order to determine whether the values fall within expected range. The foregoing is presented in attached appendix 5.2a.

Missing values were cross-checked with original documents to ascertain whether they were genuinely missing. The report, therefore, shows no error; neither were there out-of-range values, and missing values were confirmed real. Missing values were very few; below the suggested 5% upper boundary (Tabachnick & Fidell, 2007) in case(s)/variables where they occur. As fully discussed in the methodology chapter, "pairwise" deletion technique, recommended by Pallant (2007), was employed in the course of the analysis undertaken in this study. It excludes case(s) only if they are missing data required for a specific analysis, unlike "listwise" deletion approach that includes only cases with data on all the variables constituting each case, irrespective of the analysis embarked on. Consequently, the latter approach tends to lose quite a number of cases, thereby reducing drastically the total sample size. This may impact, negatively, some analyses. The former approach limits the number of cases removed. It helped eliminate or, at least, reduce any effect - if any, that may impact the sample size.

Successful conclusion of this stage, as reflected in the frequency analysis report, confirmed entered data as true version of the content of the instruments utilised. This is then followed by the verification of the validity of the questionnaire instrument through reliability test, outcome of which is discussed next.

A5.1.4 Reliability Analysis:

In the method section, the need to confirm the validity and reliability of a research instrument was discussed. In this section, the primary goal is to ascertain the appropriateness of the questionnaire instrument – builder questionnaire, that is, to ascertain its validity and reliability. Towards the identified goal, data from the studies were subjected to reliability test. Two reliability indicators, usually used, are test-retest reliability (i.e. temporal stability) and internal consistency (Pallant, 2007, 2010). The first requires administering a scale on same sample on two different occasions; thereafter, compute the correlation between the two scores obtained to determine extent of correlation. A high correlation suggests that a scale is reliable. The second indicator, internal consistency, refers to the degree to which items constituting a scale "hang together"; that is, whether they all measure same underlying construct. In this study, only the second indicator was employed due to data accessibility and time constraints. Its determination (i.e. internal consistency) for, at least, the first two studies - pilot I and II - was to inform further action(s) that may be necessary prior to the main study. Despite that, it was extended to the main study in order to explore how the scales equally behave during the latter study.

Determining the degree of internal consistency of a scale, DeVellis (2003) and Nunnally (1978) - both cited in Pallant (2007), recommended a minimum Cronbach alpha coefficient of 0.7 that a scale should attain. But values above 0.8 are preferable (Pallant, 2007), as this suggest strong reliability of a scale. Pallant (2007) further noted the sensitivity of Cronbach alpha coefficient to the number of items that makeup a scale; accordingly, short scales - with fewer than ten items – sometimes achieve Cronbach alpha value, as low as, 0.5. In such situations, the latter suggested that such scale's inter-item correlation be reported. Briggs & Cheek (1986) quoted in Pallant (2007) prescribe an optimal range for the mean inter-item correlation; this should be between 0.2 and 0.4. Equally, the reliability of a scale can vary depending on chosen sample's reactions or responses (Pallant, 2007); so, it is not out of place to observe different Cronbach alpha values for a scale administered on different samples.

In view of the discussed, a scale should be primarily declared reliable or internally consistent if at least one of the two under listed criteria is satisfied:

i. Its Cronabch alpha coefficient should be 0.7 or more, and its inter-item correlation matrix should not contain any negative value(s); or

ii. Where the Cronbach alpha coefficient is less than 0.7 and its inter-item correlation matrix does not reflect any negative value, the mean inter-item correlation should indicate a strong relationship, that is, within the range of $-1 \leq r \geq 1$, and an optimal range between items should be 0.2 to 0.4; where $r = 0$, it could be concluded that the scale is not reliable since there is no relationship between the items.

Base on the above, reliability analysis of the builder questionnaire was carried out. Detail results generated from IBM SPSS Statistics version 19 can be found in SPSS report 5.1b to 5.1g. Extracted Cronbach alpha values for the questionnaire scales, according to the phases of study, are presented in table 5.2 below. These results reflected the extent of consistency across study phases/different samples (although having same characteristics). The results achieved could not be compared with any previous one, except result of the usability scale, since they were purposely developed for this study. Hence, criteria used, to arrive at conclusion on their reliability, were based mainly on those discussed above.

On the other hand, the 19-item usability scale utilised in this study, emerged from the adaptation of 20-item questionnaire in Granic, Glavinic & Stankov (2004); had its root primarily in QUIS questionnaire that has been validated for reliability (see Chin, Diehl & Norman, 1988; Harper & Norman, 1993; Akilli, 2005). The latter questionnaire - that is, QUIS questionnaire - proved reliable, based on the results reported in Chin, Diehl & Norman (1988) thus: [i] for QUIS version 3.0 – the Cronbach alpha value was reported as 0.94 with its inter-item value varying by 0.002; [ii] QUIS version 4.0 – its reliability was reported as 0.89 with range from 0.89 to 0.90 - the small variability of the alpha value of the items was reported has indicator of high internal consistency. It thus provide basis to compare reliability result of the usability scale used in this study.

The reliability values below represent the coefficients for the seven (7) scales that constitute the instrument under consideration. They were computed and presented according to the phases of the study. Table 5.1.2 below indicates that the Cronbach alpha value for four scales – BDASUM (Builder Assimption), TUTBHV (Generated Tutor Behaviour), TUTSTRG (Tutoring Strategies), and BDRST (Builder Restriction) - were consistently above 0.7, the Cronbach alpha coefficient benchmark, for all the phases; whilst one scale – PDTIM (Production Time) – was consistently below 0.7. SPSKL (Special Skills) scale was inconsistent; it was above 0.7 in phase I and below 0.7 in phase II and main study. Equally, the three usability subscales, indicated different patterns. Subscale one – positive usability attributes – was above 0.7 criterion in all the phases, while subscales two and three – negative usability attributes I & II respectively – were below 0.7 in phase one, but greater than 0.7 in the other two phases. Detail analyses of results presented in table 5.1.2 below, with respect to each scale, follows.

Table 5.1.2: Cronbach alpha value of questionnaires scales

Scales	**No of Items**	**Study**		
		Pilot I	**Pilot II**	**Main**
Builder Assumption (BDASUM)	3	0.941	0.793	0.772
Generated Tutor Behaviour (TUTBHV)	3	0.750	0.783	0.733
Tutoring Strategies (TUTSTRG)	3	0.875	0.938	0.902
Builder Restrictions (BDRST)	3	0.875	0.812	0.848
Production Time (PDTIM)	2	0.500	0.562	0.522
Special Skills (SPSKL) scale	2	0.727	0.522	0.402
Usability (USAB) scale:				
Usability 1 (SUBUSAB1) sub-scale (+ve)	11	0.926*	0.897	0.893
Usability 2 (SUBUSAB2) sub-scale (-ve)	5	0.500	0.789	0.773
Usability 3 (SUBUSAB3) sub-scale (-ve)	3	0.273	0.770	0.805

*** pilot study I was based on 10-item, while others on 11 items (see below discussion for details)**

■ Buider Assumptions (BDASUM) scale – is a three-item scale, meant to measure extent to which the Builder prototype supports certain theoretical constructs of its underlying metamodel. Its Crobach alpha coefficient for pilot I and II were 0.941 and 0.793 respectively. These values are above the 0.7 benchmark with no negative value(s) in their inter-item correlation matrix. This informed the acceptance of the scale and its deployment in the third phase; it returned a Cronbach alpha coefficient of 0.772. Equally, its inter-item correlation matrix bears no negative value; mean inter-item correlation was 0.531 - an indication of positive correlation; and mean range between items was 0.273 – a value within the optimal range – 0.2 to 0.4 - recommended by Brigs & Cheeks(1986). Thus, the result further confirms the outcome of the reliability test conducted in study I & II. Hence, it was concluded that the scale actually measures its underlying construct. Data collected using this scale was therefore incorporated in subsequent analyses carried out in this work. See table 5.1.2 above and further details in IBM SPSS version 19 report in appendix 5.3a.

■ Generated Tutor Behaviour (TUTBHV) scale – contains three(3) items; it measures the behaviour of the tutor(s) generated from the Builder prototype, from lecturers perspectives. Based on the assumption that, if the builder system supports certain features derived from theory, tutor(s) generated from it should posses similar features. This can then be used to determine if the Builder actually achieved research objective one(1). As a measure of the scale's reliability, its Cronbach alpha coefficients for pilot I and II yielded 0.750 and 0.783 respectively (see table 5.1.2 above); their inter-item correlation matrix did not reveal any negative value(s). So, the scale was deployed for the main study; the latter returned Cronbach alpha coefficient of 0.733 (as shown above), with no negative value(s) in its inter-item correlation matrix. The latter result, hereby, confirms the reliability of the scale as measured in the first two studies. From these results, it was concluded that the scale truly measures its construct; hence, it was accepted. For detail result of the reliability test, IBM SPSS version 19 report in appendix 5.3b.

■ Tutoring Strategy (TUTSTRG) scale – is a three-item scale, meant to bring into the fore, strategies supported by the Builder prototype to accomplish the underlying theoretical construct. With respect to this scale, the two pilot studies consistently revealed Cronbach alpha coefficients that were well above the 0.7 benchmark thus: 0.875 (pilot I) and 0.938 (pilot II). This informed its deployment in the main study, which returned Cronbach alpha value of 0.902; thus, confirming the results from earlier reliability tests. For the three studies, their inter-item matrix did not reflect any negative value (see IBM SPSS version 19 report in appendix 5.3c for details).

■ Builder Restriction (BDRST) scale – has three (3) items; it measures the ability of the builder system to generate different tutors for different topics as well as their variants. In essence, it measures flexibility or non-restrictiveness of the ILABS. The piloted scale returned Cronbach alpha values of 0.875 and 0.812 for pilot I and II respectively. The scale was accepted as measuring its underlying constructs after inspecting inter-item correlation matrix of both studies and no negative values were recorded (see IBM SPSS version 19 report in appendix 5.3d). The scale was further employed in the main study; reliability results shows that a high Cronbach alpha coefficient of 0.848; no negative value(s) in its inter-item correlation matrix like previous studies. Thus, it confirmed the internal consistency of its items.

■ Production Time (PDTIM) scale – is a two-item scale, measuring time construct of the Builder prototype. From the above table 5.1.2, the scale yielded low Cronbach values of 0.500 and 0.562 for pilot I and II respectively. Although both values were below the 0.7 Cronbach alpha coefficient benchmark, the inter-item correlation matrix for each study did not reveal any negative value. Also, an inspection of the summary item statistics table (see IBM SPSS version 19 report in appendix 5.3e) shows that a strong and positive correlation exist among the items that constituted the scale thus: for pilot I, the mean inter-item correlation was 0.333 (it tallied with the minimum and maximum values, 0.333; range=0), while pilot II mean was 0.580 (equally tallied with the minimum and maximum values, 0.580, range=0). With this strong correlation achieved and mean range between items being zero(0) – a value far below the lower boundary of the optimal range suggested by Briggs & Cheeks (1986) – it thus indicates a strong relationship among items. Hence, the scale was adopted for the main study. The latter study equally returned a Cronbach alpha coefficient of 0.522, but no negative value in its inter-item correlation matrix. Its summary item statistics table revealed a mean inter-item correlation of 0.434 (a positive correlation). This value equally tallied with the minimum and maximum value reported in the latter table. Hence, the scale was accepted as a true measure of its underlying construct.

■ Special Skills (SPSKL) scale – contains two items that measures skills required from users to be able to use Builder prototype. The scale was subjected to reliability test. Cronbach alpha coefficients for pilot studies I and II are 0.727 and 0.522 respectively. The first value satisfies the reliability criteria, being above 0.7 benchmark and neither was any negative values recorded in its inter-item correlation matrix. Although, the second coefficient was below the 0.7 benchmark, reliability of the scale, with respect to the sample, was confirmed through two other criteria thus: its inter-item correlation matrix was free of negative values; the mean inter-item correlation shows a positive correlation of 0.397; and the mean range between items was zero(0) . Base on that, the scale was adopted for the main study; it returned a lower Cronbach alpha value of 0.402. Despite low value, there were no negative values in its inter-item correlation matrix, the summary item statistics table (see IBM SPSS version 19 report in appendix 5.3f) showed a positive correlation among the items with a mean inter-item correlation value of 0.293 and the mean range between items was zero(0). The result therefore confirmed the reliability of the scale.

■ Usability (USAB) scale – Unlike above discussed scales, purposely developed in their entirety for current research, the usability scale derived heavily its items from user satisfaction scales found in the literature (cf. Granic, Glavinic & Stankov, 2004; Chin, Diehl & Norman, 1988; Harper & Norman, 1993; Akilli, 2005). Instruments from those referenced research works, especially Granic, Glavinic & Stankov (2004), were adapted to suit current research. Although, Granic, Glavinic & Stankov (ibid) did not compute the Cronabch alpha coefficient in their study, they draw their instrument from the QUIS questionnaire that had been satisfied reliable. It then suffices to use the reliability result of the QUIS instrument for comparison in this study. Towards that, eighteen items relevant to this study were drawn from the 20 items scale presented in Granic, Glavinic & Stankov (ibid). Thus, it was administered as 18-item scale in pilot study I.

Thereafter, it was discovered that item 10 of the 18-item scale addressed two issues; this was later broken down into two items. So, a nineteen-item scale was administered in pilot study II and in the main study. Computation of the Cronbach alpha coefficient for pilot I and II retuned very high values of 0.894, and 0.826 respectively. However, there were negative coefficients in their inter-item matrix. An explanation that can be given is the large number of items that constitute the scale. This was resolved through factorisation, after confirming its appropriateness in this situation (Kaiser-Meyer-Olkin [KMO] measure of sampling adequacy = 0.683; Barlett's Test of Sphericity, p=0.0, was significant). Outcome of which resulted into a three-component scale; this is consistent with previous studies (Akilli, 2005; Harper & Norman, 1993; Chin, Diehl & Norman, 1988).

Reliability test conducted returned Cronbach coefficients above 0.7 for the three sub-scales and studies, except for sub-scale 2 and 3 in pilot I where their coefficients are less than 0.7. The results (see table 5.1.2 above and details in IBM SPSS version 19 report in appendix 5.3g), when compared with what is in the literature, falls within/around the coefficients achieved so far (QUIS version 4.0 => 0.89 and QUIS v3.0 => 0.94 - Chin, Diehl & Norman, 1988). Thus, the three-component solution for the usability scale was accepted as a true measure of the usability construct. Despite the above, usability construct was analysed on a per item basis (not on the whole scale or sub-scale level); noting that the sub-scales items were improperly grouped along same theme, thus making it difficult for each sub-scale to represent a specific sub-construct of the usability scale. This follow trends in the literature. For instance, previous researches that used same items in their instrument based their analysis on per item basis (see Akilli, 2005 – table 3; Harper & Norman, 1993– table 1; Chin, Diehl & Norman, 1988 - table 2).

Conclusively, all the scales were intact throughout the three-phase study, as there was no reason to modify them, except the usability scale where changes were effected after pilot I.

A5.1.5 Data Exploration / Analytical Procedure

Prior to the commencement of the quantitative analytical procedure, data collected was explored to determine relevant statistics that can be utilised. As part of the process, the following where carried out:

- Descriptive statistics were applied to data to explore its nature in order to determine statistics to use in the main/final analysis; and
- Chosen statistics were implemented accordingly.

Details of analyses employed are presented in the remaining sections of this quantitative analytical category.

■ Preliminary Analysis I - Categorical Variables

Three types of preliminary analysis were utilised in this study. The first, being analysis of categorical variables or items of the builder questionnaire instrument, A categorical item is that which has predetermined value, drawn from a fixed set of values (e.g. gender - can either be male or female). Tables 5.1.3 below, provides the spread of respondents across demographic characteristics and computer experience items – the categorical (i.e. independent) variables of the quantitative instruments utilised in the study.

An inspection of the table below reveals there were no missing values in any of the demographic items; neither were there any outliers. But, there were missing values in all the computer experience items, thus: item 1 (can you operate computer) - 1 missing; item 2 (do you know any computer programming language?) - 3 missing; item 3 (can you write codes in any programming language?) - 3 missing; and item 4 (have you used any software authoring tool before?) – 1 missing. Also, within this group, there were no outliers or out-of-range items.

In all, eighty two participants participated in the study. Their distribution according to demographic categorical variables shows: institution – 32 (39%), 25 (30.5%) and 25(30.5%) were involved from institutions Uni.B, Uni.C, and Uni.D respectively;

departments – 30 (36.6%), 28 (34.1%) and 24 (29.3%) participated from accounting, banking and finance, and other departments respectively – all within the business school of respective institutions; gender - 58 (70.7%) were male and 24(29.3%) were female; and with respect to qualification – 3 (3.7%) were university graduate, 9 (11%) were certified professionals, and 70 (85.4%) where masters degree and above holders.

Participants' distribution according to computer experience revealed another pattern thus: can you operate computer? – 81 (corresponding to 98.8%) indicated 'YES', no "NO", but one (1) missing response (1.2%); do you know any computer programming language – 48 (58.5%) stated "YES", 31 (37.8%) were "NO", and others missing (3.7%); can you write computer codes – 24 (29.3%) answered "YES", 55 (67.1%) answered "No", and 3 (3.7%) were missing; and the item, "have you used any authoring tool before" – 44 (53.7%) were "Yes", 37 (45.1%) were "No", and one(1) response, equivalent to 1.2% , was missing.

With the above described patterns, the effect of categorical variables on continuous variables was only considered for few variables that had good spread of participants, such as institution, department, gender, do know any computer programming language, and have you used any authoring tool before. Their choice was based on the spread of participants' responses across each variable's valid groups. It aligns with Stevens (1996, p.249) cited in Pallant (2010, p.207) suggestion, that the sample size of a categorical variable groups should be reasonably similar (e.g. largest/smallest should not exceed 1.5). This is necessary if one is applying parametric statistics. Therefore, reading through the below table, it is quite obvious that above identified variables satisfy this criterion.

While respondents are almost evenly distributed across the segments of each variable, the below table 5.1.3 reveals that 98.8% of respondents can operate computer; this implies, almost all respondents are computer literate, hence may not have any technical hindrance towards evaluation of the ILABS. Also, a large number (85.4%) had postgraduate degree, at least, a master's degree; as well, majority (36.6%) are from accounting department. These two factors show the level of competence and relevance which may equally provide a good platform for the evaluation of the tool under consideration.

Table 5.1.3: Demographic/Computer Experience Characteristics of Respondents

Categorical Variables	Frequency (N)	Percentage (%)

Demographic variables:			
University:			
Uni.B		32	39.0
Uni.C		25	30.5
Uni.D		25	30.5
Department:			
Accounting		30	36.6
Banking & finance		28	34.1
Others		24	29.3
Gender:			
Male		58	70.7
Female		24	29.3
Qualification:			
University graduate		3	3.7
Professional		9	11.0
Postgraduate		70	85.3
Others		---	---
Computer Experience:			
1. Can you operate computer?	YES	81	98.8
	NO	---	---
	Missing	1	1.2
2. Do you know any computer programming language?	YES	48	58.5
	NO	33	37.8
	Missing	3	3.7
3. Can you write codes in any computer programming lanaguge?	YES	24	29.3
	NO	55	67.1
	Missing	3	3.7
4. Have you used any software authoring tool before?	YES	44	53.7
	NO	37	45.1
	Missing	1	1.2

■ Preliminary Analysis II - Continuous Variables

A continuous variable can be defined as item with changing value - not fixed; and may (or may not) derive its value from other variable(s). It could be dependent or independent variable. Unlike categorical variables, usually, preliminary analysis of continuous variable employ descriptive statistics such as mean, median, standard deviation, skewness, kurtosis etc. These statistics enable the exploration of the nature of data. As well, it can indicate the normality status of a distribution via the skewness and kurtosis values.

In this category, respective builder questionnaire scales constitute the continuous variables in this study. They were therefore subjected to descriptive statistics (see appendix 5.2b), since frequency analysis was not appropriate (or, at least, sufficient) to explore the data; neither was it appropriate to reveal its nature, especially, the normality of the data. The latter information is required to ascertain the appropriateness of parametric statistics. Table 5.1.4 below provide descriptive statistics for the builder questionnaire instrument continuous variables. Unlike the categorical variables, the continuous variables presented in the table below did not reflect any missing value. This can be attributed to the measures taken during data collection, in which maximum cooperation was sought with participants and it was impressed on them, the need to complete all sections diligently; this eventually pay-off with respect to this section of the research instrument.

Table 5.4: Descriptive statistics for Builder Quesionnaire Continuous Variables (or Scales)

		BDASUM	TUTBHV	TUTSTRG	BDRST	PDTIM	SPSKL	USAB
N	Valid	82	82	82	82	82	82	82
	Missing	0	0	0	0	0	0	0
Mean		4.5447	4.52439	4.56098	4.38211	3.84756	4.28049	3.96406
5% Trimmed mean		4.54968	4.52710	4.56775	4.41057	3.85772	4.33198	3.96876
Median		4.66667	4.66667	4.66667	4.33333	4.0000	4.5000	3.89474
Variance		0.169	0.156	0.208	0.338	0.634	0.420	0.127
Standard deviation		0.410801	0.395553	0.456257	0.581191	0.796168	0.648348	0.355802
Minimum		4.0	4.0	4.0	3.0	2.0	2.5	3.158
Maximum		5.0	5.0	5.0	5.0	5.0	5.0	4.579
Skewness		-0.155	-0.110	-0.202	-0.388	0.066	-0.904	-0.104
Kurtosis		-1.592	-1.500	-1.828	-1.061	-1.122	0.695	-0.817

The table also reflected the minimum and maximum scores for each variable/scale; as well, the mean, 5% trimmed mean, median and standard deviation for respective scale was presented. PDTIM scale had the largest deviation (0.796 approx.) and USAB scale had the lowest (0.355 approx.). Median was 4.0 or above for the scales except for USAB scale with a lower median (3.895 approx). When compared to mean, median was greater than mean except for two scales, thus: BDRST(mean=4.382 , median=4.333) and USAB(mean=3.964, median=3.895). What this comparison suggests, is that, more scores tends towards the highest score of the scale (i.e. 5) than to the lowest score (i.e. 1). This provides insight into the pattern of responses to each scale, suggesting agreement to constructs measured. This is fully examined in main analysis later. The mean and 5% trimmed mean of each scale were very close, an indication that extreme scores did not influence any of the scale's mean (Pallant, 2007).

Two other important items were given, the skewness and kurtosis values. They provide information on the distribution of scores on the continuous variables (i.e. the scales). The skewness and kurtosis signify or indicate extent of normality of data under consideration. For a perfectly normal distribution, one would expect the skewness and the kurtosis values to be zero[0] (Pallant, 2010). A normal distribution is one which has the greatest frequency of scores in the middle, with smaller frequencies towards the extremes. But the result from table above shows negative skewness for all the scales except "PDTIM" scale – with positive value. The negatively skewed scales indicate that scores are clustered at the high end of the normality graph - the right-hand side of a graph; while for the positive skewed scale - "PDTIM" scale – point to a distribution with scores clustered at the low end of the graph - the left-hand side of a normal graph. Also, kurtosis had negative values in six (6) scales and positive value in one(1) scale. The results, therefore, indicates that all the scales, except "SPSKL" with positive kurtosis, are having relatively flat distributions; that is, too many cases are in the extremes, so their graphs are not peaked at the centre. While the "SPSKL" with positive kurtosis, suggests a distribution that is rather peaked (i.e. clustered in the centre). The skewness and kurtosis results of the scales, therefore, suggest lack of normality. However, it is inconclusive as further confirmation would be required (see normality section below).

■ Preliminary Analysis III – Normality Assessment

Normality of a distribution, describes a symmetrical, bell-shape curve, *"which has the greatest frequency of scores in the middle with smaller frequencies towards the extremes"* (Gravetter & Wallnau, 2004, p. 48 – cited in Pallant, 2010, p. 59). To some extent, normality of a distribution can be determined by the skewness and kurtosis measures; although, confirmatory evidence may still be necessary through the conduct of a normality test. Such test is usually assessed by Kolmogorov-Smirnov statistic. When undertaken, a value greater than 0.05 indicates non-significance (i.e. sig. > 0.05), implying normality; whilst a value below indicates significance, that is, absence of normality. Also, histogram, Normal Q-Q plot and boxplot could help inform the normality of a distribution (Pallant, 2007).

As part of the preliminary analysis process, data collected was subjected to normality test using Kolmogorov-Smirnov statistic. Table 5.1,5 below shows the result of the test for all the scales. Using the benchmark stated above, sig. > 0.05, all the scales were significant (i.e. Sig. < 0.05 for all the scales). Thus, it denotes that the scales are non-normally distributed; further, it confirms the results of the skewness and kurtosis. The implication, therefore, is that parametric statistics may likely not apply. But, Pallant

(2007) noted that the robustness of parametric statistics can accommodate non-normal distribution. Also, the latter noted the possibility of some scales and measures, either positively or negatively skewed, yet not having any associated problem.

Table 5.1.5: Tests of Normality for Builder Questionnaire Data

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
Builder Assumption scale	.232	82	.000	.811	82	.000
Generated Tutor Behaviour scale	.190	82	.000	.834	82	.000
Tutoring strategy scale	.320	82	.000	.725	82	.000
Builder Restriction scale	.222	82	.000	.862	82	.000
Production Time scale	.161	82	.000	.904	82	.000
Special Skill scale	.223	82	.000	.843	82	.000
Usability scale	.102	82	.036	.970	82	.055
Positive Usability Attributes sub-scale	.116	82	.008	.958	82	.009
Negative Usability Attributes I sub-scale	.108	82	.019	.961	82	.013
Negative Usability Attributes II sub-scale	.203	82	.000	.899	82	.000

However, the results and aforementioned arguments, throws-up three routes that can be explored, thus: [i] to use non-parametric statistics; [ii] to hold on to parametric statistics using the data as it is; or [iii] to transform data to normality, then apply parametric statistics. Initially, the last option was totally ruled out for fear of biasing the result of the analysis. But, later, it was explored by trying each of the three most common transformation technique – square root, log and inverse transformation techniques. None of these techniques succeeded in normalising the data, thereby limiting selection to the first two options. In order to determine which of the options to take, the study took cognisance of several arguments, for and against the use of parametric statistics for non-normal distribution (sees Pallant, 2010; Tabachnick & Fidell, 2007; Kerr, Hall & Kozub, 2002; Foster, 2001). Consequently, the study opted for parametric statistics, unless otherwise stated. The decision was based on evidence in the literature that points to the robustness of parametric test to accommodate non-compliance with some assumptions, including normality assumption (see Kerr, Hall & Kozub, 2002, p.54; Foster, 2001, p.17; Pallant, 2010). Equally, it relies on the power of parametric statistics to detect small significance, wherever they exist.

Appendix 5.2a: Frequency Distribution of Demographic & Computer Experience Factors of Builder Questionnaire

Frequencies

Statistics

		Institution	Department	Gender	Highest Qualification	Can you operate computer	Do you know any computer programming language	Can you write computer codes	Have you used any authoring tool before
N	Valid	82	82	82	82	81	79	79	81
	Missing	0	0	0	0	1	3	3	1
Minimum		2	1	1	1	1	1	1	1
Maximum		4	3	2	3	1	2	2	2

Frequency Table

Institution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2 Uni.B	32	39.0	39.0	39.0
	3 Uni.C	25	30.5	30.5	69.5
	4 Uni.D	25	30.5	30.5	100.0
	Total	82	100.0	100.0	

Department

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Accounting	30	36.6	36.6	36.6
	2 Banking & Finance	28	34.1	34.1	70.7
	3 Others	24	29.3	29.3	100.0
	Total	82	100.0	100.0	

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Male	58	70.7	70.7	70.7
	2 Female	24	29.3	29.3	100.0
	Total	82	100.0	100.0	

Highest Qualification

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 University Graduate	3	3.7	3.7	3.7
	2 Professional	9	11.0	11.0	14.6
	3 Postgraduate	70	85.4	85.4	100.0
	Total	82	100.0	100.0	

Can you operate computer

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	81	98.8	100.0	100.0
Missing	System	1	1.2		
Total		82	100.0		

Do you know any computer programming language

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	48	58.5	60.8	60.8
	2 No	31	37.8	39.2	100.0
	Total	79	96.3	100.0	
Missing	System	3	3.7		
Total		82	100.0		

Can you write computer codes

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	24	29.3	30.4	30.4
	2 No	55	67.1	69.6	100.0
	Total	79	96.3	100.0	
Missing	System	3	3.7		
Total		82	100.0		

Have you used any authoring tool before

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	44	53.7	54.3	54.3
	2 No	37	45.1	45.7	100.0
	Total	81	98.8	100.0	
Missing	System	1	1.2		
Total		82	100.0		

Appendix 5.2b: Descriptive Statistics of Builder Questionnaire Scales

Statistics

		Builder Assumption scale	Generated Tutor Behaviour scale	Tutoring strategy scale	Builder Restriction scale	Production Time scale	Special Skill scale	Usability scale
N	Valid	82	82	82	82	82	82	82
	Missing	0	0	0	0	0	0	0
Mean		4.54472	4.52439	4.56098	4.38211	3.84756	4.28049	3.96406
Median		4.66667	4.66667	4.66667	4.33333	4.00000	4.50000	3.89474
Std. Deviation		.410801	.395553	.456257	.581191	.796168	.648348	.355802
Skewness		-.155	-.110	-.202	-.388	.066	-.904	-.104
Std. Error of Skewness		.266	.266	.266	.266	.266	.266	.266
Kurtosis		-1.592	-1.500	-1.828	-1.061	-1.122	.695	-.817
Std. Error of Kurtosis		.526	.526	.526	.526	.526	.526	.526
Minimum		4.000	4.000	4.000	3.000	2.000	2.500	3.158
Maximum		5.000	5.000	5.000	5.000	5.000	5.000	4.579

Study Type: Pilot Study I & II

Case Processing Summary

Type of Study			N	%
1 Pilot study I	Cases	Valid	4	100.0
		Excluded[a]	0	.0
		Total	4	100.0
2 Pilot study II	Cases	Valid	24	100.0
		Excluded[a]	0	.0
		Total	24	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Type of Study	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
1 Pilot study I	.941	.956	3
2 Pilot study II	.793	.791	3

Item Statistics

Type of Study		Mean	Std. Deviation	N
1 Pilot study I	bdasum1	4.25	.500	4
	bdasum2	4.00	.816	4
	bdasum3	4.00	.816	4
2 Pilot study II	bdasum1	4.67	.482	24
	bdasum2	4.50	.511	24
	bdasum3	4.38	.495	24

Inter-Item Correlation Matrix

Type of Study		bdasum1	bdasum2	bdasum3
1 Pilot study I	bdasum1	1.000	.816	.816
	bdasum2	.816	1.000	1.000
	bdasum3	.816	1.000	1.000
2 Pilot study II	bdasum1	1.000	.707	.365
	bdasum2	.707	1.000	.602
	bdasum3	.365	.602	1.000

Summary Item Statistics

Type of Study		Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
1 Pilot study I	Inter-Item Correlations	.878	.816	1.000	.184	1.225	.009	3
2 Pilot study II	Inter-Item Correlations	.558	.365	.707	.342	1.936	.025	3

Item-Total Statistics

Type of Study		Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
1 Pilot study I	bdasum1	8.00	2.667	.816	.	1.000
	bdasum2	8.25	1.583	.973	.	.842
	bdasum3	8.25	1.583	.973	.	.842
2 Pilot study II	bdasum1	8.88	.810	.602	.506	.752
	bdasum2	9.04	.650	.792	.637	.535
	bdasum3	9.17	.841	.527	.370	.828

Scale Statistics

Type of Study	Mean	Variance	Std. Deviation	N of Items
1 Pilot study I	12.25	4.250	2.062	3
2 Pilot study II	13.54	1.563	1.250	3

Study Type: Main Study (Builder Assumption Scale)

Case Processing Summary

		N	%
Cases	Valid	82	100.0
	Excluded[a]	0	.0
	Total	82	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.772	.772	3

Item Statistics

	Mean	Std. Deviation	N
bdasum1	4.65	.481	82
bdasum2	4.50	.503	82
bdasum3	4.49	.503	82

Inter-Item Correlation Matrix

	bdasum1	bdasum2	bdasum3
bdasum1	1.000	.689	.416
bdasum2	.689	1.000	.488
bdasum3	.416	.488	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.531	.416	.689	.273	1.657	.016	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
bdasum1	8.99	.753	.640	.483	.656
bdasum2	9.13	.685	.697	.523	.587
bdasum3	9.15	.818	.493	.250	.815

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
13.63	1.519	1.232	3

Appendix 5.3b: Reliability Analysis of Generated Tutor Behaviour Scale

Study Type: Pilot Study I & II

Case Processing Summary

Type of Study			N	%
1 Pilot study I	Cases	Valid	4	100.0
		Excluded[a]	0	.0
		Total	4	100.0
2 Pilot study II	Cases	Valid	24	100.0
		Excluded[a]	0	.0
		Total	24	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Type of Study	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
1 Pilot study I	.750	.747	3
2 Pilot study II	.783	.784	3

Item Statistics

Type of Study		Mean	Std. Deviation	N
1 Pilot study I	tutbhv1	4.25	.500	4
	tutbhv2	4.75	.500	4
	tutbhv3	4.50	.577	4
2 Pilot study II	tutbhv1	4.33	.482	24
	tutbhv2	4.58	.504	24
	tutbhv3	4.33	.482	24

Inter-Item Correlation Matrix

Type of Study		tutbhv1	tutbhv2	tutbhv3
1 Pilot study I	tutbhv1	1.000	.333	.577
	tutbhv2	.333	1.000	.577
	tutbhv3	.577	.577	1.000
2 Pilot study II	tutbhv1	1.000	.598	.625
	tutbhv2	.598	1.000	.418
	tutbhv3	.625	.418	1.000

Summary Item Statistics

Type of Study		Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
1 Pilot study I	Inter-Item Correlations	.496	.333	.577	.244	1.732	.016	3
2 Pilot study II	Inter-Item Correlations	.547	.418	.625	.207	1.494	.010	3

Item-Total Statistics

Type of Study		Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
1 Pilot study I	tutbhv1	9.25	.917	.522	.333	.727
	tutbhv2	8.75	.917	.522	.333	.727
	tutbhv3	9.00	.667	.707	.500	.500
2 Pilot study II	tutbhv1	8.92	.688	.725	.528	.589
	tutbhv2	8.67	.754	.564	.360	.769
	tutbhv3	8.92	.775	.581	.394	.748

Scale Statistics

Type of Study	Mean	Variance	Std. Deviation	N of Items
1 Pilot study I	13.50	1.667	1.291	3
2 Pilot study II	13.25	1.500	1.225	3

Study Type: Main Study

Case Processing Summary

		N	%
Cases	Valid	82	100.0
	Excluded[a]	0	.0
	Total	82	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.733	.733	3

Item Statistics

	Mean	Std. Deviation	N
tutbhv1	4.40	.493	82
tutbhv2	4.67	.473	82
tutbhv3	4.50	.503	82

Inter-Item Correlation Matrix

	tutbhv1	tutbhv2	tutbhv3
tutbhv1	1.000	.469	.522
tutbhv2	.469	1.000	.441
tutbhv3	.522	.441	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.478	.441	.522	.081	1.184	.001	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
tutbhv1	9.17	.687	.585	.344	.611
tutbhv2	8.90	.756	.522	.273	.686
tutbhv3	9.07	.686	.563	.322	.638

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
13.57	1.408	1.187	3

Study Type: Pilot Study I & II

Case Processing Summary

Type of Study			N	%
1 Pilot study I	Cases	Valid	4	100.0
		Excluded[a]	0	.0
		Total	4	100.0
2 Pilot study II	Cases	Valid	24	100.0
		Excluded[a]	0	.0
		Total	24	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Type of Study	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
1 Pilot study I	.875	.884	3
2 Pilot study II	.938	.938	3

Item Statistics

Type of Study		Mean	Std. Deviation	N
1 Pilot study I	tutstrg1	4.75	.500	4
	tutstrg2	4.75	.500	4
	tutstrg3	4.50	.577	4
2 Pilot study II	tutstrg1	4.58	.504	24
	tutstrg2	4.63	.495	24
	tutstrg3	4.54	.509	24

Inter-Item Correlation Matrix

Type of Study		tutstrg1	tutstrg2	tutstrg3
1 Pilot study I	tutstrg1	1.000	1.000	.577
	tutstrg2	1.000	1.000	.577
	tutstrg3	.577	.577	1.000
2 Pilot study II	tutstrg1	1.000	.742	.919
	tutstrg2	.742	1.000	.842
	tutstrg3	.919	.842	1.000

Summary Item Statistics

Type of Study		Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
1 Pilot study I	Inter-Item Correlations	.718	.577	1.000	.423	1.732	.048	3
2 Pilot study II	Inter-Item Correlations	.834	.742	.919	.177	1.238	.006	3

Item-Total Statistics

Type of Study		Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
1 Pilot study I	tutstrg1	9.25	.917	.870	.	.727
	tutstrg2	9.25	.917	.870	.	.727
	tutstrg3	9.50	1.000	.577	.	1.000
2 Pilot study II	tutstrg1	9.17	.928	.867	.848	.914
	tutstrg2	9.13	.984	.809	.716	.958
	tutstrg3	9.21	.868	.944	.901	.852

Scale Statistics

Type of Study	Mean	Variance	Std. Deviation	N of Items
1 Pilot study I	14.00	2.000	1.414	3
2 Pilot study II	13.75	2.022	1.422	3

Study Type: Main Study

Case Processing Summary

		N	%
Cases	Valid	82	100.0
	Excluded[a]	0	.0
	Total	82	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.902	.902	3

Item Statistics

	Mean	Std. Deviation	N
tutstrg1	4.56	.499	82
tutstrg2	4.59	.496	82
tutstrg3	4.54	.502	82

Inter-Item Correlation Matrix

	tutstrg1	tutstrg2	tutstrg3
tutstrg1	1.000	.702	.804
tutstrg2	.702	1.000	.757
tutstrg3	.804	.757	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.754	.702	.804	.102	1.145	.002	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
tutstrg1	9.12	.874	.804	.667	.861
tutstrg2	9.10	.904	.768	.597	.891
tutstrg3	9.15	.843	.846	.719	.825

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
13.68	1.874	1.369	3

Study Type: Pilot Study I & II

Case Processing Summary

Type of Study			N	%
1 Pilot study I	Cases	Valid	4	100.0
		Excluded[a]	0	.0
		Total	4	100.0
2 Pilot study II	Cases	Valid	24	100.0
		Excluded[a]	0	.0
		Total	24	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Type of Study	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
1 Pilot study I	.875	.884	3
2 Pilot study II	.812	.820	3

Item Statistics

Type of Study		Mean	Std. Deviation	N
1 Pilot study I	bdrst1	4.75	.500	4
	bdrst2	4.75	.500	4
	bdrst3	4.50	.577	4
2 Pilot study II	bdrst1	4.33	.482	24
	bdrst2	4.42	.654	24
	bdrst3	4.13	.741	24

Inter-Item Correlation Matrix

Type of Study		bdrst1	bdrst2	bdrst3
1 Pilot study I	bdrst1	1.000	1.000	.577
	bdrst2	1.000	1.000	.577
	bdrst3	.577	.577	1.000
2 Pilot study II	bdrst1	1.000	.506	.609
	bdrst2	.506	1.000	.696
	bdrst3	.609	.696	1.000

Summary Item Statistics

Type of Study		Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
1 Pilot study I	Inter-Item Correlations	.718	.577	1.000	.423	1.732	.048	3
2 Pilot study II	Inter-Item Correlations	.604	.506	.696	.189	1.374	.007	3

Item-Total Statistics

Type of Study		Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
1 Pilot study I	bdrst1	9.25	.917	.870	.	.727
	bdrst2	9.25	.917	.870	.	.727
	bdrst3	9.50	1.000	.577	.	1.000
2 Pilot study II	bdrst1	8.54	1.650	.609	.384	.817
	bdrst2	8.46	1.216	.689	.495	.715
	bdrst3	8.75	.978	.756	.573	.652

Scale Statistics

Type of Study	Mean	Variance	Std. Deviation	N of Items
1 Pilot study I	14.00	2.000	1.414	3
2 Pilot study II	12.88	2.636	1.624	3

Study Type: Main Study

Case Processing Summary

		N	%
Cases	Valid	82	100.0
	Excluded[a]	0	.0
	Total	82	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.848	.851	3

Item Statistics

	Mean	Std. Deviation	N
bdrst1	4.37	.639	82
bdrst2	4.52	.593	82
bdrst3	4.26	.750	82

Inter-Item Correlation Matrix

	bdrst1	bdrst2	bdrst3
bdrst1	1.000	.596	.678
bdrst2	.596	1.000	.694
bdrst3	.678	.694	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.656	.596	.694	.098	1.164	.002	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
bdrst1	8.78	1.531	.697	.490	.806
bdrst2	8.62	1.621	.708	.510	.802
bdrst3	8.89	1.210	.767	.590	.746

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
13.15	3.040	1.744	3

Study Type: Pilot Study

Case Processing Summary

Type of Study			N	%
1 Pilot study I	Cases	Valid	4	100.0
		Excluded[a]	0	.0
		Total	4	100.0
2 Pilot study II	Cases	Valid	24	100.0
		Excluded[a]	0	.0
		Total	24	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Type of Study	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
1 Pilot study I	.500	.500	2
2 Pilot study II	.562	.734	2

Item Statistics

Type of Study		Mean	Std. Deviation	N
1 Pilot study I	pdtim1	3.75	.500	4
	pdtim2	2.25	.500	4
2 Pilot study II	pdtim1	4.33	.482	24
	pdtim2	3.38	1.245	24

Inter-Item Correlation Matrix

Type of Study		pdtim1	pdtim2
1 Pilot study I	pdtim1	1.000	.333
	pdtim2	.333	1.000
2 Pilot study II	pdtim1	1.000	.580
	pdtim2	.580	1.000

Summary Item Statistics

Type of Study		Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
1 Pilot study I	Inter-Item Correlations	.333	.333	.333	.000	1.000	.000	2
2 Pilot study II	Inter-Item Correlations	.580	.580	.580	.000	1.000	.000	2

Item-Total Statistics

Type of Study		Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
1 Pilot study I	pdtim1	2.25	.250	.333	.111	.
	pdtim2	3.75	.250	.333	.111	.
2 Pilot study II	pdtim1	3.38	1.549	.580	.337	.
	pdtim2	4.33	.232	.580	.337	.

Scale Statistics

Type of Study	Mean	Variance	Std. Deviation	N of Items
1 Pilot study I	6.00	.667	.816	2
2 Pilot study II	7.71	2.476	1.574	2

Study Type: Main Study

Case Processing Summary

		N	%
Cases	Valid	82	100.0
	Excluded[a]	0	.0
	Total	82	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.522	.606	2

Item Statistics

	Mean	Std. Deviation	N
pdtim1	4.26	.625	82
pdtim2	3.44	1.218	82

Inter-Item Correlation Matrix

	pdtim1	pdtim2
pdtim1	1.000	.434
pdtim2	.434	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.434	.434	.434	.000	1.000	.000	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
pdtim1	3.44	1.484	.434	.189	.
pdtim2	4.26	.390	.434	.189	.

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
7.70	2.536	1.592	2

Study Type: Pilot Study

Case Processing Summary

Type of Study			N	%
1 Pilot study I	Cases	Valid	4	100.0
		Excluded[a]	0	.0
		Total	4	100.0
2 Pilot study II	Cases	Valid	24	100.0
		Excluded[a]	0	.0
		Total	24	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Type of Study	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
1 Pilot study I	.727	.732	2
2 Pilot study II	.522	.568	2

Item Statistics

Type of Study		Mean	Std. Deviation	N
1 Pilot study I	spskl1rvs	4.50	.577	4
	spskl2	4.75	.500	4
2 Pilot study II	spskl1rvs	4.17	.963	24
	spskl2	4.46	.588	24

Inter-Item Correlation Matrix

Type of Study		spskl1rvs	spskl2
1 Pilot study I	spskl1rvs	1.000	.577
	spskl2	.577	1.000
2 Pilot study II	spskl1rvs	1.000	.397
	spskl2	.397	1.000

Summary Item Statistics

Type of Study		Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
1 Pilot study I	Inter-Item Correlations	.577	.577	.577	.000	1.000	.000	2
2 Pilot study II	Inter-Item Correlations	.397	.397	.397	.000	1.000	.000	2

Item-Total Statistics

Type of Study		Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
1 Pilot study I	spskl1rvs	4.75	.250	.577	.333	.
	spskl2	4.50	.333	.577	.333	.
2 Pilot study II	spskl1rvs	4.46	.346	.397	.157	.
	spskl2	4.17	.928	.397	.157	.

Scale Statistics

Type of Study	Mean	Variance	Std. Deviation	N of Items
1 Pilot study I	9.25	.917	.957	2
2 Pilot study II	8.63	1.723	1.313	2

Study Type: Main Study

Case Processing Summary

		N	%
Cases	Valid	82	100.0
	Excluded[a]	0	.0
	Total	82	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.402	.453	2

Item Statistics

	Mean	Std. Deviation	N
spskl1rvs	4.09	1.009	82
spskl2	4.48	.571	82

Inter-Item Correlation Matrix

	spskl1rvs	spskl2
spskl1rvs	1.000	.293
spskl2	.293	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.293	.293	.293	.000	1.000	.000	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
spskl1rvs	4.48	.327	.293	.086	.
spskl2	4.09	1.017	.293	.086	.

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
8.56	1.681	1.297	2

Study Type: Main Study (Usability sub-Scale I - Positive Items)

Case Processing Summary

		N	%
Cases	Valid	82	100.0
	Excluded[a]	0	.0
	Total	82	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.893	.901	11

Item Statistics

	Mean	Std. Deviation	N
usab4 - Builder can be described as interesting	4.35	.575	82
usab5 - Builder can be described as stimulating	4.37	.533	82
usab7 - Builder cab be described as usable	4.37	.639	82
usab9 - I feel in control when I am using the system	3.96	.728	82
usab10 - Builder system uses terms that is understandable	4.01	.923	82
usab11 - Builder system uses terms that is familiar to me	3.63	.794	82
usab13 - It is easy to understand the objects on the Builder system's interface	4.15	.569	82
usab15 - I get what I expect when I click on objects on the Builder system interface	3.90	.730	82
usab17 - I feel efficient when using the Builder system	4.16	.555	82
usab18 - Builder system can be characterised as innovative	4.55	.570	82
usab19 - Overall, I am satisfied with the Builder system	4.49	.593	82

Inter-Item Correlation Matrix

	usab4 - Builder can be described as interesting	usab5 - Builder can be described as stimulating	usab7 - Builder cab be described as usable	usab9 - I feel in control when I am using the system	usab10 - Builder system uses terms that is understandable	usab11 - Builder system uses terms that is familiar to me	usab13 - It is easy to understand the objects on the Builder system's interface	usab15 - I get what I expect when I click on objects on the Builder system interface	usab17 - I feel efficient when using the Builder system	usab18 - Builder system can be characterised as innovative	usab19 - Overall, I am satisfied with the Builder system
usab4 - Builder can be described as interesting	1.000	.338	.484	.386	.341	.260	.142	.083	.287	.380	.466
usab5 - Builder can be described as stimulating	.338	1.000	.762	.544	.643	.262	.716	.410	.636	.631	.717
usab7 - Builder cab be described as usable	.484	.762	1.000	.561	.600	.292	.530	.342	.601	.629	.762
usab9 - I feel in control when I am using the system	.386	.544	.561	1.000	.773	.532	.401	.481	.626	.347	.585
usab10 - Builder system uses terms that is understandable	.341	.643	.600	.773	1.000	.444	.584	.423	.430	.292	.643
usab11 - Builder system uses terms that is familiar to me	.260	.262	.292	.532	.444	1.000	.038	.278	.133	.149	.410
usab13 - It is easy to understand the objects on the Builder system's interface	.142	.716	.530	.401	.584	.038	1.000	.391	.512	.549	.517
usab15 - I get what I expect when I click on objects on the Builder system interface	.083	.410	.342	.481	.423	.278	.391	1.000	.496	.101	.225
usab17 - I feel efficient when using the Builder system	.287	.636	.601	.626	.430	.133	.512	.496	1.000	.541	.512
usab18 - Builder system can be characterised as innovative	.380	.631	.629	.347	.292	.149	.549	.101	.541	1.000	.623
usab19 - Overall, I am satisfied with the Builder system	.466	.717	.762	.585	.643	.410	.517	.225	.512	.623	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.452	.038	.773	.735	20.341	.034	11

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
usab4 - Builder can be described as interesting	41.59	23.209	.433	.323	.894
usab5 - Builder can be described as stimulating	41.57	21.680	.800	.764	.876
usab7 - Builder cab be described as usable	41.57	20.964	.780	.724	.874
usab9 - I feel in control when I am using the system	41.98	20.345	.771	.784	.874
usab10 - Builder system uses terms that is understandable	41.93	19.056	.748	.793	.877
usab11 - Builder system uses terms that is familiar to me	42.30	22.289	.402	.460	.899
usab13 - It is easy to understand the objects on the Builder system's interface	41.79	22.364	.603	.679	.885
usab15 - I get what I expect when I click on objects on the Builder system interface	42.04	22.283	.452	.447	.895
usab17 - I feel efficient when using the Builder system	41.78	22.173	.661	.681	.882
usab18 - Builder system can be characterised as innovative	41.39	22.587	.558	.644	.887
usab19 - Overall, I am satisfied with the Builder system	41.45	21.337	.775	.718	.876

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
45.94	25.935	5.093	11

Study Type: Main Study (Usability sub-Scale II - Negative Items)

Case Processing Summary

		N	%
Cases	Valid	82	100.0
	Excluded[a]	0	.0
	Total	82	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.773	.799	5

Item Statistics

	Mean	Std. Deviation	N
usab1rvs - Builder can be described as annoying	4.27	.668	82
usab2rvs - Builder can be described as confusing	4.04	.693	82
usab3rvs - Builder can be described as frustrating	4.11	.667	82
usab12rvs - Builder system needs more introductory explanations	2.07	.716	82
usab16rvs - It is difficult to move around the Builder system	3.28	.959	82

Inter-Item Correlation Matrix

	usab1rvs - Builder can be described as annoying	usab2rvs - Builder can be described as confusing	usab3rvs - Builder can be described as frustrating	usab12rvs - Builder system needs more introductory explanations	usab16rvs - It is difficult to move around the Builder system
usab1rvs - Builder can be described as annoying	1.000	.726	.821	.320	.363
usab2rvs - Builder can be described as confusing	.726	1.000	.793	.318	.244
usab3rvs - Builder can be described as frustrating	.821	.793	1.000	.319	.318
usab12rvs - Builder system needs more introductory explanations	.320	.318	.319	1.000	.203
usab16rvs - It is difficult to move around the Builder system	.363	.244	.318	.203	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.442	.203	.821	.617	4.036	.054	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
usab1rvs - Builder can be described as annoying	13.50	4.722	.749	.702	.667
usab2rvs - Builder can be described as confusing	13.73	4.816	.672	.650	.690
usab3rvs - Builder can be described as frustrating	13.66	4.721	.750	.757	.667
usab12rvs - Builder system needs more introductory explanations	15.70	5.622	.355	.129	.790
usab16rvs - It is difficult to move around the Builder system	14.49	4.944	.346	.145	.825

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
17.77	7.341	2.709	5

Study Type: Main Study (Usability sub-Scale III - Negative Items)

Case Processing Summary

		N	%
Cases	Valid	82	100.0
	Excluded[a]	0	.0
	Total	82	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.803	.805	3

Item Statistics

	Mean	Std. Deviation	N
usab6rvs - Builder can be described as tiresome	3.76	.695	82
usab8rvs - Builder can be described as unpleasant	4.06	.635	82
usab14rvs - Builder system is slow	3.79	.698	82

Inter-Item Correlation Matrix

	usab6rvs - Builder can be described as tiresome	usab8rvs - Builder can be described as unpleasant	usab14rvs - Builder system is slow
usab6rvs - Builder can be described as tiresome	1.000	.649	.531
usab8rvs - Builder can be described as unpleasant	.649	1.000	.558
usab14rvs - Builder system is slow	.531	.558	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.579	.531	.649	.119	1.223	.003	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
usab6rvs - Builder can be described as tiresome	7.85	1.386	.665	.463	.714
usab8rvs - Builder can be described as unpleasant	7.55	1.485	.690	.485	.693
usab14rvs - Builder system is slow	7.82	1.460	.598	.360	.785

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
11.61	2.957	1.720	3

T-Test

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Builder Assumption scale	1 Male	58	4.59195	.400015	.052525
	2 Female	24	4.43056	.422543	.086251
Generated Tutor Behaviour scale	1 Male	58	4.54598	.383369	.050339
	2 Female	24	4.47222	.427516	.087266
Tutoring strategy scale	1 Male	58	4.56897	.450292	.059126
	2 Female	24	4.54167	.479659	.097910
Builder Restriction scale	1 Male	58	4.40805	.621396	.081593
	2 Female	24	4.31944	.476290	.097222
Production Time scale	1 Male	58	4.03448	.782832	.102791
	2 Female	24	3.39583	.642332	.131115
Special Skill scale	1 Male	58	4.26724	.695950	.091383
	2 Female	24	4.31250	.527762	.107729

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Builder Assumption scale	Equal variances assumed	.149	.700	1.635	80	.106	.161398	.098691	-.035002	.357799
	Equal variances not assumed			1.598	40.950	.118	.161398	.100986	-.042554	.365351
Generated Tutor Behaviour scale	Equal variances assumed	1.442	.233	.766	80	.446	.073755	.096250	-.117789	.265299
	Equal variances not assumed			.732	39.106	.468	.073755	.100744	-.130002	.277512
Tutoring strategy scale	Equal variances assumed	1.100	.297	.245	80	.807	.027299	.111386	-.194367	.248964
	Equal variances not assumed			.239	40.652	.813	.027299	.114378	-.203752	.258349
Builder Restriction scale	Equal variances assumed	5.113	.026	.626	80	.533	.088602	.141594	-.193179	.370382
	Equal variances not assumed			.698	55.666	.488	.088602	.126924	-.165691	.342894
Production Time scale	Equal variances assumed	2.934	.091	3.531	80	.001	.638649	.180857	.278733	.998566
	Equal variances not assumed			3.833	52.030	.000	.638649	.166605	.304337	.972962
Special Skill scale	Equal variances assumed	1.486	.226	-.286	80	.776	-.045259	.158260	-.360206	.269689
	Equal variances not assumed			-.320	56.255	.750	-.045259	.141267	-.328222	.237705

T-Test

Group Statistics

	Have you used any authoring tool before	N	Mean	Std. Deviation	Std. Error Mean
Builder Assumption scale	1 Yes	44	4.43939	.417775	.062982
	2 No	37	4.68468	.359577	.059114
Generated Tutor Behaviour scale	1 Yes	44	4.49242	.376909	.056821
	2 No	37	4.57658	.413123	.067917
Tutoring strategy scale	1 Yes	44	4.63636	.459289	.069240
	2 No	37	4.47748	.448182	.073681
Builder Restriction scale	1 Yes	44	4.42424	.544475	.082083
	2 No	37	4.34234	.630924	.103723
Production Time scale	1 Yes	44	3.56818	.728098	.109765
	2 No	37	4.17568	.765726	.125885
Special Skill scale	1 Yes	44	4.42045	.505127	.076151
	2 No	37	4.10811	.764990	.125764

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Builder Assumption scale	Equal variances assumed	3.382	.070	-2.803	79	.006	-.245291	.087511	-.419477	-.071104
	Equal variances not assumed			-2.840	78.949	.006	-.245291	.086378	-.417224	-.073357
Generated Tutor Behaviour scale	Equal variances assumed	1.080	.302	-.958	79	.341	-.084152	.087845	-.259004	.090699
	Equal variances not assumed			-.950	73.774	.345	-.084152	.088551	-.260604	.092299
Tutoring strategy scale	Equal variances assumed	.214	.645	1.568	79	.121	.158886	.101326	-.042798	.360571
	Equal variances not assumed			1.571	77.233	.120	.158886	.101109	-.042439	.360211
Builder Restriction scale	Equal variances assumed	2.478	.119	.627	79	.532	.081900	.130590	-.178032	.341832
	Equal variances not assumed			.619	71.675	.538	.081900	.132273	-.181801	.345601
Production Time scale	Equal variances assumed	.508	.478	-3.653	79	.000	-.607494	.166284	-.938475	-.276513
	Equal variances not assumed			-3.637	75.172	.001	-.607494	.167019	-.940200	-.274788
Special Skill scale	Equal variances assumed	3.591	.062	2.199	79	.031	.312346	.142050	.029602	.595091
	Equal variances not assumed			2.124	60.436	.038	.312346	.147022	.018303	.606390

Appendix 5.5: Spearman Correlation of Builder Assumption & Generated Tutor Behaviour Scales

Nonparametric Correlations

			Builder Assumption scale	bdasum1	bdasum2	bdasum3	Generated Tutor Behaviour scale	tutbhv1	tutbhv2	tutbhv3
Spearman's rho	Builder Assumption scale	Correlation Coefficient	1.000	.828**	.869**	.785**	.542**	.507**	.423**	.377**
		Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.000
		N	82	82	82	82	82	82	82	82
	bdasum1	Correlation Coefficient	.828**	1.000	.689**	.416**	.553**	.555**	.404**	.383**
		Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.000	.000
		N	82	82	82	82	82	82	82	82
	bdasum2	Correlation Coefficient	.869**	.689**	1.000	.488**	.452**	.423**	.337**	.317**
		Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.002	.004
		N	82	82	82	82	82	82	82	82
	bdasum3	Correlation Coefficient	.785**	.416**	.488**	1.000	.353**	.294**	.320**	.244*
		Sig. (2-tailed)	.000	.000	.000	.	.001	.007	.003	.027
		N	82	82	82	82	82	82	82	82
	Generated Tutor Behaviour scale	Correlation Coefficient	.542**	.553**	.452**	.353**	1.000	.832**	.772**	.815**
		Sig. (2-tailed)	.000	.000	.000	.001	.	.000	.000	.000
		N	82	82	82	82	82	82	82	82
	tutbhv1	Correlation Coefficient	.507**	.555**	.423**	.294**	.832**	1.000	.469**	.522**
		Sig. (2-tailed)	.000	.000	.000	.007	.000	.	.000	.000
		N	82	82	82	82	82	82	82	82
	tutbhv2	Correlation Coefficient	.423**	.404**	.337**	.320**	.772**	.469**	1.000	.441**
		Sig. (2-tailed)	.000	.000	.002	.003	.000	.000	.	.000
		N	82	82	82	82	82	82	82	82
	tutbhv3	Correlation Coefficient	.377**	.383**	.317**	.244*	.815**	.522**	.441**	1.000
		Sig. (2-tailed)	.000	.000	.004	.027	.000	.000	.000	.
		N	82	82	82	82	82	82	82	82

**. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

Appendix 5.6: Frequency Distribution of Production Scale Items

Statistics

		pdtim1	pdtim2
N	Valid	82	82
	Missing	0	0

Frequency Table

pdtim1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2 Disagress	2	2.4	2.4	2.4
	3 Neither agree nor disagree	2	2.4	2.4	4.9
	4 Agree	51	62.2	62.2	67.1
	5 Strongly Agree	27	32.9	32.9	100.0
	Total	82	100.0	100.0	

pdtim2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly disagree	1	1.2	1.2	1.2
	2 Disagress	26	31.7	31.7	32.9
	3 Neither agree nor disagree	12	14.6	14.6	47.6
	4 Agree	22	26.8	26.8	74.4
	5 Strongly Agree	21	25.6	25.6	100.0
	Total	82	100.0	100.0	

Frequencies

Statistics

		Usability scale	usab1rvs	usab2rvs	usab3rvs	usab6rvs	usab8rvs	usab12rvs	usab14rvs	usab16rvs	usab4	usab5	usab7	usab9	usab10	usab11	usab13	usab15	usab17	usab18	usab19
N	Valid	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82
	Missing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mean		3.96406	4.27	4.04	4.11	3.76	4.06	2.07	3.79	3.28	4.35	4.37	4.37	3.96	4.01	3.63	4.15	3.90	4.16	4.55	4.49
Median		3.89474	4.00	4.00	4.00	4.00	4.00	2.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	5.00	5.00
Std. Dev.		.355802	.668	.693	.667	.695	.635	.716	.698	.959	.575	.533	.639	.728	.923	.794	.569	.730	.555	.570	.593
Min.		3.158	2	2	3	2	2	1	2	1	3	3	3	2	2	2	3	2	2	3	3
Max.		4.579	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5

T-Test

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Usability scale	1 Male	58	4.00726	.324210	.042571
	2 Female	24	3.85965	.411407	.083978

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Usability scale	Equal variances assumed	1.370	.245	1.730	80	.087	.147610	.085313	-.022168	.317389
	Equal variances not assumed			1.568	35.397	.126	.147610	.094152	-.043452	.338673

T-Test

Group Statistics

	Have you used any authoring tool before	N	Mean	Std. Deviation	Std. Error Mean
Usability scale	1 Yes	44	4.00239	.369025	.055633
	2 No	37	3.92745	.340494	.055977

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Usability scale	Equal variances assumed	.421	.519	.943	79	.349	.074939	.079477	-.083256	.233133
	Equal variances not assumed			.950	78.291	.345	.074939	.078920	-.082170	.232047

A6.1.1 Data Collection/Phases

This appendix augment chapter 6 of this thesis. It discusses the preliminary analysis of eTutor questionnaire data only, which addresses research question two. The research instrument for this aspect of the study was mainly questionnaire. It was administered in three phases. Two phases were pilot studies and the third phase was the main study. The pilot studies were undertaken purposely to test the research instrument in order to evaluate its reliability and validity.

The target subjects, for this aspect of the research, were mainly students from higher education institutions. Table 6.1.1 below shows the responses with respect to the three data collection phases and the higher education providing institutions that participated. The code naming of institutions, as reflected in the table below, was done to enforce the confidentiality of the data collected.

Table 6.1.1: Number of Participants in Pilot and Main Studies

Phases	Institution	Samples
Pilot I	Uni.A	13
Pilot II	Uni.B	15
	Uni.C	15
	Total$_{\text{pilot II}}$	**30**
Main Study	Uni.B	151
	Uni.C	149
Total$_{\text{main study}}$		**300**

Data collected was subjected to various statistical analyses. Preliminary analysis was undertaken to explore the study's data and guide the analytical procedure adopted. Thereafter, descriptive and bivariate statistical analyses were utilised to reveal findings from the main study's data.

A6.1.2 Data Screening/Cleaning

Prior to analysis, data collected for this aspect of the research was cleaned, as recommended in Pallant (2007, 2010), by checking for data entry errors and outliers. The process was implemented by generating frequency statistics of all variables, which showed frequency of valid cases, missing values, and minimum and maximum values.

This statistics was used to determine whether the values fall within expected range (see appendix 6.2a). Missing values were cross-checked with original documents to ascertain whether they were genuinely missing. The report did not show any error, neither were there outliers, and missing values were confirmed real. Thus, data entered into IBM SPSS Statistics version 19 was assumed as true version of the content of the administered questionnaires.

A6.1.3 Reliability Analysis

Similar to research question one, data collected with respect to the research question two were analysed to determine the reliability of the eTutor questionnaire scales. See appendix 6.3a to 6.3g for the IBM SPSS version 19 reliability analysis reports for the seven(7) scales considered. Extracted Cronbach alpha values from the reliability analysis reports, for the eTutor questionnaire scales, are presented in table 6.1.2 below. From the table, the Cronbach alpha values for scales - CPVSB, RFDBK, LNEFTV and CVSBLN - were consistently above 0.7 for all the phases, whilst two scales - TIMFDBK and MISCP - were consistently below 0.7. Only CCVSB scale was inconsistent; it was above 0.7 in phase I and below 0.7 in phases II and III.

Table 6.2: Cronbach alpha values for eTutor Questionnaires scales

Scales	No of Items	Phase		
		I	**II**	**III**
Cognitive Process Visibility (CPVSB)	7	0.857	0.886	0.809
Conversation aids Cognitive Visibility (CCVSB)	2	0.776	0.476	0.609
Timely Feedback (TIMFDBK)	2	0.689	0.514	0.277
Relevant Feedback (RELFDBK)	6*	0.780	0.763	0.726
Misconception (MISCP)	3*	0.542	0.640	0.630
Learning Effectiveness (LNEFTV)	12*	0.931	0.950	0.855
Cognitive Visibility & Learning (CVSBLN)	5*	0.849	0.879	0.839

*** Items that hang-up to measure respective construct and consistent throughout the three phases**

But, after thorough analysis of the reliability analysis results of the scales (as indicated in the table 6.1.2 above), they were taken to be true measures of their respective constructs. Also, as shown in the table 6.1.2 above, it could be observed that scales with higher number of items yielded higher Cronbach alpha values. This further confirms Pallant(2007) claim that Cronbach alpha values are sensitive to scales with high number of items. However, one or more items of some scales (i.e. RELFDBK, MISCP, LNEFTV and CVSBLN)s were dropped because they did not hang-up with others in their respective scale. This was undertaken in line with the argument of Pallant (2010); she suggested removal of items with negative inter-item correlation values and whose removal could improve the Cronbach alpha value of respective scales (as indicated by the "*Cronbach alpha value if deleted*" column of the item-total statistics table of the reliability analysis report for each scale). In that regard, scales with "*" as shown in the above table, thus indicates that number of items from the eTutor questionnaire instrument that was considered for each respective scale.

A6.1.4 Preliminary Analysis

Prior to the main analytical procedure, data collected was explored. This involved the application of descriptive statistics to explore the nature of the data, in order to determine appropriate statistics that can be employed to reveal the research findings from the data set. Two levels of descriptive statistics were employed:

- First, descriptive analysis of categorical variables of the questionnaire. These are questionnaire items with predetermined value drawn from a fixed set of values (e.g. gender - can be male or female). In category are the demographic and computer experience variables. At this level, frequency analysis was employed.
- Second, descriptive analysis of the continuous variables of the questionnaire. A continuous variable can be defined as item with changing value (not fixed; may or may not derive its value from other variable[s]; and could be dependent or independent variable[s]). The statistics employed in this case includes the mean, standard deviation, normal test etc.

■ Preliminary Analysis I – Categorical Variables

Table 6.1.3 below presents the frequency analysis of the demographic characteristics and computer experience of respondents as extracted from appendix 6.2a. Thus, the table indicated that analysis of users perceptions can be undertaken across institutions, departments (accounting versus other departments), gender, qualification (O/L versus Others), and eTutoring experience, since this variables had almost evenly distributed responses.

Respondents' distribution across computer experience variables shows that 99% can operate computer. Hence, it was assumed that students would not have any technical problem that may hinder the use of the eTutor. Also, 90.3% like learning via computer;

this should aid a well informed view about the tutoring system being evaluated. Thus, for the latter two variables, there was no basis for comparison of users' views since responses were not evenly distributed.

Also, table 6.1.3 above shows that missing values were insignificant compared to the sample size of the data. Hence, for statistical analysis based on variable(s) with missing value(s), a pair-wise deletion approach was applied. It ensures that only case(s) with missing value(s) is/are removed from such analysis, with respect to the concerned variable where missing value(s) occur.

Therefore, relating the result back to data sources (that is, completed questionnaires) helped confirm the accuracy of the data entered into the statistical package. Also, it helped identify whether there were returned instruments with incomplete response; a situation where some items in an instrument are not addressed by respondent(s). It also reflects the distribution of the responses/participation according to the demographic dimensions and computer experience of respective instruments.

Table 6.1.3: Frequency Analysis of Demographic/Computer Experience Characteristics

Categorical Variables		Main Study	
		Frequency (N)	**Percentage (%)**
Demographic variables:			
University:			
Uni.B		151	50.3
Uni.C		149	49.7
Department:			
Accounting		178	59.3
Other Departments		122	40.7
Gender:			
Male		164	54.7
Female		134	44.7
[Missing]		2	0.7
Age:			
16 – 25		195	65.0
26 – 35		95	31.7
36 - 45		7	2.3
46 and above		2	0.7
Others		---	---
Missing		1	0.3
Qualification:			
GCE/WAEC/NECO/GCSE		101	33.7
Diloma		174	58.0
A-Level		10	3.3
Others		13	4.3
Missing		2	0.7
Computer Experience:			
1. Can you operate computer?	YES	297	99.0
	NO	3	1.0
2. Do you enjoy using computer at all?	YES	297	99.0
	NO	3	1.0
3. Do you know any computer programming language?	YES	167	55.7
	NO	129	43.0
	Missing	4	1.3
4. Can you write codes in any computer programming lanaguge?	YES	87	29.0
	NO	209	69.7
	Missing	4	1.3
5. Do you like learning via computer?	YES	271	90.3
	NO	27	9.0
	Missing	2	0.7
6. Have you used any eTutor software before now?	YES	147	49.0
	NO	149	49.7
	Missing	4	1.3
7. If question 6 is YES, was it in accounting and/or finance related subject?	YES	81	27.0
	NO	134	44.7
	Not Applicable	80	26.7
	Missing	5	1.7

■ Preliminary Analysis II - Continuous Variables

For this category, the eTutor questionnaire scales were subjected to different descriptive statistics. These variables constitute the continuous (independent and/or dependent variables of the instruments. For these set of variables, frequency analysis was considered inappropriate or sufficient to explore the data; neither is it appropriate to reveal its nature - its normality, a basic requirement that must be ascertained before any parametric statistics can be applied. Thus, they were subjected to other forms of descriptive analysis, aimed at revealing the mean, standard deviation, and assessing normality of the data set (see appendix 6.2b).

Table 6.1.4 below provides information on the eTutor data set (that was extracted from appendix 6.2b); it shows the distribution of respondents' responses across constructs/scales measured. One such vital information provided is the missing values statistics which revealed thus: RFDBK (1 missing), LNEFTV (8 missing) and CVSBLN (4 missing). The missing responses were traced back to their original documents and they were confirmed actually missing. The above outcome further buttresses the fact that it may be humanly impossible to prevent missing responses in a research involving human subjects. Therefore, allowances must be made to accommodate them by reducing or, possibly, eliminating their effect on data analysis. On this note, "pairwise" deletion approach was adopted, considering its advantages over "listwise" deletion technique, which include the reduction of deleted cases. (Pallant, 2010).

The table 6.1.4 below also presented among others, the mean, 5% trimmed mean, minimum and maximum scores, and standard deviation of scores - with the lowest being 0.538 (for LNEFTV scale) and the highest being 0.718 (CCVSB scale). The difference between mean and 5% trimmed mean is an indication of the effect an extreme score can have on the mean of a scale. However, the result shows a narrow gap between the mean and 5% trimmed mean of each respective scale; hence, it could be concluded that extreme scores did not impact the mean score of each of the scales. Assume they do, they may impact analysis; consequentially, it may be required to verify further - whether they are real or caused by data entry error. But, in this case, such did not occur.

Two other items were presented in table 6.1.4, the skewness and kurtosis values. They represent the distribution of scores on the continuous variables (or scales). The values presented in the table above, therefore, indicated negative skewness for all the scales. Thus implies that scores were clustered towards the high end of respective scale's normality graph. This suggests scores were non-normally distributed on the scales. For the kurtosis, four scales had negative values, indicating a flatten shape graph. This is a situation of many extreme scores. Other scales were positive, an indication of rather peaked graph (with scores clustered in the centre). For normality, the skewness and kurtosis should be zero(0). Hence, both set of values revealed lack of normality in the data. This conclusion was further investigated through normality assessment, which was discussed below.

Table6.1.4: Descriptive statistics for Continuous Variables (or Scale) in eTutor Data

		CPVSB	CCVSB	TIMFDBK	RELFDBK	MISCP	LNEFTV	CVSBLN
N	Valid	286	297	292	289	297	280	296
	Missing	14	3	8	11	3	20	4
Mean		3.727	3.895	3.628	3.874	3.680	3.905	4.048
5% Trimmed mean		3.752	3.930	3.641	3.877	3.696	3.917	4.075
Median		3.714	4.000	3.500	3.833	3.667	3.917	4.000
Variance		0.416	0.515	0.478	0.305	0.457	0.289	0.386
Standard deviation		0.645	0.718	0.691	0.552	0.676	0.538	0.621
Minimum		1.714	1.500	1.500	2.500	1.667	2.500	2.000
Maximum		5.000	5.000	5.000	5.000	5.000	5.000	5.000
Skewness		-0.373	-0.590	-0.232	-0.162	-0.222	-0.252	-0.616
Kurtosis		0.037	0.209	-0.059	-0.579	-0.125	-0.436	0.423

■ Preliminary Analysis III – Normality Assessment

Although, skewness and kurtosis do measure to some extent the normality of a distribution; for confirmatory evidence, actual normality test should be conducted. As part of the descriptive statistics, data from various instruments were subjected to normality test. Usually, normality of a distribution can be assessed by Kolmogorov-Smirnov statistic (Park, 2008; Chan, 2003; Pallant, 2007). A non-significant value (sig.), that is, when sig. > 0.05, indicates normality; but a value below indicates significance – that is, absence of normality.

Table 6.1.5: Tests of Normality for eTutor Data

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
CPVSB scale	.082	286	.000	.958	286	.000
CCVSB Scale	.215	297	.000	.907	297	.000
TIMFDBK Scale	.156	292	.000	.943	292	.000
RELFDBK Scale	.093	289	.000	.945	289	.000
MISCP Scale	.123	297	.000	.965	297	.000
LNEFTV Scale	.075	280	.001	.949	280	.000
CVSBLN Scale	.161	296	.000	.915	296	.000

Table 6.1.5 above shows the normality values for the questionnaire scales. Each of the scales had sig. Value that is below the 0.05 normality benchmark. This signifies that sigma was significant for all the scales; an indication that scores on each scale was not normally distributed. Thus, it confirms further the non-normality prediction from the skewness values presented in table 6.1.4 above. The implication is that parametric statistics may likely not apply. But, Pallant (2007) noted that parametric statistics can accommodate non-normal distribution due to its robustness. Also, the latter noted the possibility of some scales/measures not having any associated problem, despite being positively or negatively skewed. As a result, three options are open for consideration, similar to options stated in previous chapter, thus:

- [i] to use non-parametric statistics;
- [ii] to hold on to parametric statistics using the data as it is; or
- [iii] to transform data to normality, then apply parametric statistics.

However, the third option was ruled out, in order to preserve the data integrity and to avoid any bias that data transformation may introduce into the analysis. Thus were left with the first two options. But, based on earlier argument in favour of parametric statistics' robustness, it was therefore considered.

Frequencies

Statistics

		Institution	Department	Gender	Age	Highest Qualification	cmpexp1	cmpexp2	cmpexp3	cmpexp4	cmpexp5	cmpexp6	cmpexp7	cmpexp8
N	Valid	300	300	299	299	299	300	300	296	296	299	300	300	0
	Missing	0	0	1	1	1	0	0	4	4	1	0	0	300

Frequency Table

Institution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2 Uni.B	151	50.3	50.3	50.3
	3 Uni.C	149	49.7	49.7	100.0
	Total	300	100.0	100.0	

Department

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Accounting	178	59.3	59.3	59.3
	2 Banking & Finance	14	4.7	4.7	64.0
	3 Acturial Science and Insurance	77	25.7	25.7	89.7
	4 Economics	7	2.3	2.3	92.0
	5 Others	24	8.0	8.0	100.0
	Total	300	100.0	100.0	

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Male	164	54.7	54.8	54.8
	2 Female	135	45.0	45.2	100.0
	Total	299	99.7	100.0	
Missing	System	1	.3		
Total		300	100.0		

Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 16 – 25	195	65.0	65.2	65.2
	2 26 – 35	95	31.7	31.8	97.0
	3 36 – 45	7	2.3	2.3	99.3
	4 46 and above	2	.7	.7	100.0
	Total	299	99.7	100.0	
Missing	System	1	.3		
Total		300	100.0		

Highest Qualification

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 GCE/WASC/NECO/GCSE	101	33.7	33.8	33.8
	2 Diploma	174	58.0	58.2	92.0
	3 A-Levels	10	3.3	3.3	95.3
	4 Others	14	4.7	4.7	100.0
	Total	299	99.7	100.0	
Missing	System	1	.3		
Total		300	100.0		

cmpexp1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	297	99.0	99.0	99.0
	2 No	3	1.0	1.0	100.0
	Total	300	100.0	100.0	

cmpexp2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	297	99.0	99.0	99.0
	2 No	3	1.0	1.0	100.0
	Total	300	100.0	100.0	

cmpexp3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	167	55.7	56.4	56.4
	2 No	129	43.0	43.6	100.0
	Total	296	98.7	100.0	
Missing	System	4	1.3		
Total		300	100.0		

cmpexp4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	87	29.0	29.4	29.4
	2 No	209	69.7	70.6	100.0
	Total	296	98.7	100.0	
Missing	System	4	1.3		
Total		300	100.0		

cmpexp5

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	272	90.7	91.0	91.0
	2 No	27	9.0	9.0	100.0
	Total	299	99.7	100.0	
Missing	System	1	.3		
Total		300	100.0		

cmpexp6

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	147	49.0	49.0	49.0
	2 No	153	51.0	51.0	100.0
	Total	300	100.0	100.0	

cmpexp7

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	81	27.0	27.0	27.0
	2 No	139	46.3	46.3	73.3
	3 Not Applicable	80	26.7	26.7	100.0
	Total	300	100.0	100.0	

cmpexp8

		Frequency	Percent
Missing	System	300	100.0

Appendix 6.2b: Descriptive Statistics of eTutor Questionnaire Scales

Descriptive Statistics

Statistics

		Cognitive Process Visibility Scale	Conversation Aid Cognitive Visibility Scale	Timely Feedback scale	Relevant Feedback scale	Misconception scale	Learning Effectiveness scale	Cognitive Visibility & Learning scale
N	Valid	300	300	300	299	300	292	296
	Missing	0	0	0	1	0	8	4
Mean		3.72714	3.89500	3.62833	3.87402	3.68000	3.90582	4.04797
Median		3.71429	4.00000	3.50000	3.83333	3.66667	3.91667	4.00000
Std. Deviation		.645073	.717536	.691102	.552195	.675944	.537694	.621204
Skewness		-.373	-.590	-.232	-.162	-.222	-.252	-.616
Std. Error of Skewness		.141	.141	.141	.141	.141	.143	.142
Kurtosis		.037	.209	-.059	-.579	-.125	-.436	.423
Std. Error of Kurtosis		.281	.281	.281	.281	.281	.284	.282

Study Type: Main study only

Case Processing Summary

		N	%
Cases	Valid	300	100.0
	Excluded[a]	0	.0
	Total	300	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.809	.809	7

Item Statistics

	Mean	Std. Deviation	N
cpvsb1	3.74	.935	300
cpvsb2	3.75	.892	300
cpvsb3	3.72	.947	300
cpvsb4	3.69	.951	300
cpvsb5	3.66	.980	300
cpvsb6	3.70	1.024	300
cpvsb7	3.83	.879	300

Inter-Item Correlation Matrix

	cpvsb1	cpvsb2	cpvsb3	cpvsb4	cpvsb5	cpvsb6	cpvsb7
cpvsb1	1.000	.425	.328	.406	.412	.334	.239
cpvsb2	.425	1.000	.410	.390	.302	.361	.321
cpvsb3	.328	.410	1.000	.524	.403	.327	.388
cpvsb4	.406	.390	.524	1.000	.495	.431	.263
cpvsb5	.412	.302	.403	.495	1.000	.420	.296
cpvsb6	.334	.361	.327	.431	.420	1.000	.428
cpvsb7	.239	.321	.388	.263	.296	.428	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.376	.239	.524	.285	2.193	.005	7

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
cpvsb1	22.35	15.719	.512	.296	.789
cpvsb2	22.34	15.843	.529	.306	.787
cpvsb3	22.37	15.250	.574	.380	.779
cpvsb4	22.40	14.957	.615	.428	.771
cpvsb5	22.43	15.122	.565	.350	.780
cpvsb6	22.39	14.955	.554	.341	.782
cpvsb7	22.26	16.355	.459	.268	.798

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
26.09	20.390	4.516	7

Appendix 6.3b: Reliability Analysis of Conversation aid Cognitive Visibility Scale

Study Type: Main study only

Case Processing Summary

		N	%
Cases	Valid	300	100.0
	Excluded[a]	0	.0
	Total	300	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.609	.611	2

Item Statistics

	Mean	Std. Deviation	N
ccvsb1	3.89	.802	300
ccvsb2	3.90	.889	300

Inter-Item Correlation Matrix

	ccvsb1	ccvsb2
ccvsb1	1.000	.440
ccvsb2	.440	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.440	.440	.440	.000	1.000	.000	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
ccvsb1	3.90	.790	.440	.194	.
ccvsb2	3.89	.643	.440	.194	.

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
7.79	2.059	1.435	2

Appendix 6.3c: Reliability Analysis of Timely Feedback Scale

Study Type: Main study only

Case Processing Summary

		N	%
Cases	Valid	300	100.0
	Excluded[a]	0	.0
	Total	300	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.277	.279	2

Item Statistics

	Mean	Std. Deviation	N
timfdbk1	3.89	.847	300
timfdbk2rvs	3.36	.963	300

Inter-Item Correlation Matrix

	timfdbk1	timfdbk2rvs
timfdbk1	1.000	.162
timfdbk2rvs	.162	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.162	.162	.162	.000	1.000	.000	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
timfdbk1	3.36	.927	.162	.026	.
timfdbk2rvs	3.89	.718	.162	.026	.

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
7.25	1.909	1.382	2

Appendix 6.3d: Reliability Analysis of Relevant Feedback Scale

Study Type: Main study only

Case Processing Summary

		N	%
Cases	Valid	299	99.7
	Excluded[a]	1	.3
	Total	300	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.726	.733	6

Item Statistics

	Mean	Std. Deviation	N
relfdbk1	4.11	.706	299
relfdbk2	3.79	.850	299
relfdbk3	4.06	.764	299
relfdbk5	3.60	.894	299
relfdbk6	3.61	1.022	299
relfdbk7	4.08	.828	299

Inter-Item Correlation Matrix

	relfdbk1	relfdbk2	relfdbk3	relfdbk5	relfdbk6	relfdbk7
relfdbk1	1.000	.406	.404	.079	.249	.416
relfdbk2	.406	1.000	.284	.366	.334	.305
relfdbk3	.404	.284	1.000	.205	.226	.576
relfdbk5	.079	.366	.205	1.000	.449	.180
relfdbk6	.249	.334	.226	.449	1.000	.232
relfdbk7	.416	.305	.576	.180	.232	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.314	.079	.576	.497	7.246	.015	6

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
relfdbk1	19.14	8.609	.451	.311	.693
relfdbk2	19.45	7.832	.509	.296	.674
relfdbk3	19.18	8.229	.494	.376	.680
relfdbk5	19.65	8.155	.396	.280	.708
relfdbk6	19.64	7.426	.450	.263	.697
relfdbk7	19.16	7.990	.492	.382	.679

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
23.24	10.977	3.313	6

Study Type: Main study only

Case Processing Summary

		N	%
Cases	Valid	300	100.0
	Excluded[a]	0	.0
	Total	300	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.630	.631	3

Item Statistics

	Mean	Std. Deviation	N
miscp1	3.72	.896	300
miscp2	3.52	.883	300
miscp3	3.80	.896	300

Inter-Item Correlation Matrix

	miscp1	miscp2	miscp3
miscp1	1.000	.376	.272
miscp2	.376	1.000	.440
miscp3	.272	.440	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.363	.272	.440	.167	1.614	.006	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
miscp1	7.32	2.277	.382	.156	.611
miscp2	7.52	2.043	.511	.264	.428
miscp3	7.24	2.176	.429	.207	.546

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
11.04	4.112	2.028	3

Appendix 6.3f: Reliability Analysis of Learning Effectiveness Scale

Study Type: Main study only

Case Processing Summary

		N	%
Cases	Valid	292	97.3
	Excluded[a]	8	2.7
	Total	300	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.855	.859	12

Item Statistics

	Mean	Std. Deviation	N
lneftv1	4.25	.694	292
lneftv2	4.14	.776	292
lneftv3	3.93	.919	292
lneftv4	3.98	.889	292
lneftv5	3.88	.837	292
lneftv6	4.01	.869	292
lneftv8	3.54	.999	292
lneftv9	3.83	.800	292
lneftv10	3.45	.985	292
lneftv11	3.96	.822	292
lneftv12	3.95	.895	292
lneftv13	3.95	.866	292

Inter-Item Correlation Matrix

	lneftv1	lneftv2	lneftv3	lneftv4	lneftv5	lneftv6	lneftv8	lneftv9	lneftv10	lneftv11	lneftv12	lneftv13
lneftv1	1.000	.707	.451	.398	.339	.373	.039	.233	-.010	.275	.329	.336
lneftv2	.707	1.000	.543	.517	.454	.426	.074	.305	.084	.434	.386	.435
lneftv3	.451	.543	1.000	.385	.428	.383	.072	.139	.157	.324	.351	.397
lneftv4	.398	.517	.385	1.000	.569	.565	.152	.266	.262	.347	.400	.418
lneftv5	.339	.454	.428	.569	1.000	.563	.236	.262	.313	.393	.368	.456
lneftv6	.373	.426	.383	.565	.563	1.000	.142	.279	.213	.414	.446	.434
lneftv8	.039	.074	.072	.152	.236	.142	1.000	.214	.626	.285	.114	.187
lneftv9	.233	.305	.139	.266	.262	.279	.214	1.000	.313	.298	.348	.359
lneftv10	-.010	.084	.157	.262	.313	.213	.626	.313	1.000	.322	.114	.265
lneftv11	.275	.434	.324	.347	.393	.414	.285	.298	.322	1.000	.464	.533
lneftv12	.329	.386	.351	.400	.368	.446	.114	.348	.114	.464	1.000	.577
lneftv13	.336	.435	.397	.418	.456	.434	.187	.359	.265	.533	.577	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.338	-.010	.707	.718	-68.474	.022	12

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
lneftv1	42.62	37.053	.485	.527	.847
lneftv2	42.73	35.318	.619	.640	.838
lneftv3	42.93	35.216	.511	.382	.845
lneftv4	42.89	34.393	.619	.482	.837
lneftv5	42.99	34.605	.642	.477	.836
lneftv6	42.86	34.607	.613	.462	.838
lneftv8	43.33	36.853	.312	.410	.861
lneftv9	43.04	36.812	.431	.253	.850
lneftv10	43.42	36.025	.392	.497	.854
lneftv11	42.91	35.136	.597	.420	.839
lneftv12	42.92	34.942	.556	.438	.842
lneftv13	42.92	34.375	.641	.486	.836

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
46.87	41.632	6.452	12

Appendix 6.3g: Reliability Analysis of Cognitive Visibility & Learning Scale

Study Type: Main study only

Case Processing Summary

		N	%
Cases	Valid	309	98.7
	Excluded[a]	4	1.3
	Total	313	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.839	.839	5

Item Statistics

	Mean	Std. Deviation	N
cvsbln1	4.13	.715	309
cvsbln2	4.08	.781	309
cvsbln3	4.04	.840	309
cvsbln4	4.02	.828	309
cvsbln5	4.00	.779	309

Inter-Item Correlation Matrix

	cvsbln1	cvsbln2	cvsbln3	cvsbln4	cvsbln5
cvsbln1	1.000	.503	.386	.501	.419
cvsbln2	.503	1.000	.564	.566	.469
cvsbln3	.386	.564	1.000	.643	.491
cvsbln4	.501	.566	.643	1.000	.554
cvsbln5	.419	.469	.491	.554	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items

Case Processing Summary

		N	%
Cases	Valid	309	98.7
	Excluded[a]	4	1.3
	Total	313	100.0

Inter-Item Correlations	.510	.386	.643	.258	1.668	.005	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
cvsbln1	16.14	6.902	.555	.335	.829
cvsbln2	16.19	6.285	.665	.450	.801
cvsbln3	16.24	6.051	.663	.485	.802
cvsbln4	16.26	5.881	.731	.545	.781
cvsbln5	16.27	6.497	.603	.370	.818

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
20.28	9.499	3.082	5

Appendix 6.4: Main Effect of Institution Factor on Cognitive Process Visibility & Conversation Aid Cognitive Visibility Scales

T-Test

Group Statistics

	Institution	N	Mean	Std. Deviation	Std. Error Mean
Cognitive Process Visibility Scale	2 Uni.B	151	3.95364	.621070	.050542
	3 Uni.C	149	3.49760	.586811	.048073
Conversation Aid Cognitive Visibility Scale	2 Uni.B	151	3.97682	.617354	.050240
	3 Uni.C	149	3.81208	.800077	.065545

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Cognitive Process Visibility Scale	Equal variances assumed	1.571	.211	6.535	298	.000	.456039	.069780	.318715	.593363
	Equal variances not assumed			6.538	297.442	.000	.456039	.069753	.318766	.593312
Conversation Aid Cognitive Visibility Scale	Equal variances assumed	8.369	.004	1.998	298	.047	.164741	.082444	.002494	.326987
	Equal variances not assumed			1.995	278.231	.047	.164741	.082584	.002172	.327310

Appendix 6.5: Main Effect of Department Factor on Cognitive Process Visibility & Conversation Aid Cognitive Visibility Scales

T-Test

Group Statistics

	Collapsed Departments	N	Mean	Std. Deviation	Std. Error Mean
Cognitive Process Visibility Scale	1 Accounting	178	3.78892	.680849	.051032
	2 Other Depts	122	3.63700	.579968	.052508
Conversation Aid Cognitive Visibility Scale	1 Accounting	178	4.01404	.654040	.049022
	2 Other Depts	122	3.72131	.771521	.069850

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Cognitive Process Visibility Scale	Equal variances assumed	5.341	.022	2.014	298	.045	.151922	.075435	.003470	.300375
	Equal variances not assumed			2.075	284.201	.039	.151922	.073221	.007798	.296046
Conversation Aid Cognitive Visibility Scale	Equal variances assumed	8.263	.004	3.537	298	.000	.292733	.082758	.129869	.455598
	Equal variances not assumed			3.430	231.207	.001	.292733	.085336	.124598	.460869

Appendix 6.6: Main and Interaction Effects of Institution, Department & Gender Factors on Timely Feedback Scale

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
Institution	2	Uni.B	151
	3	Uni.C	148
Collapsed Departments	1	Accounting	178
	2	Other Depts	121
Gender	1	Male	164
	2	Female	135

Descriptive Statistics

Dependent Variable:Timely Feedback scale

Institution	Collapsed Departments	Gender	Mean	Std. Deviation	N
2 Uni.B	1 Accounting	1 Male	3.53968	.789557	63
		2 Female	3.86585	.622740	41
		Total	3.66827	.742636	104
	2 Other Depts	1 Male	3.55556	.683608	18
		2 Female	3.51724	.700545	29
		Total	3.53191	.686866	47
	Total	1 Male	3.54321	.763207	81
		2 Female	3.72143	.673708	70
		Total	3.62583	.726220	151
3 Uni.C	1 Accounting	1 Male	3.58824	.621223	34
		2 Female	3.70000	.597001	40
		Total	3.64865	.606640	74
	2 Other Depts	1 Male	3.59184	.689757	49
		2 Female	3.68000	.748331	25
		Total	3.62162	.706190	74
	Total	1 Male	3.59036	.658641	83
		2 Female	3.69231	.653669	65
		Total	3.63514	.656197	148
Total	1 Accounting	1 Male	3.55670	.731993	97
		2 Female	3.78395	.612057	81
		Total	3.66011	.687609	178
	2 Other Depts	1 Male	3.58209	.683113	67
		2 Female	3.59259	.720810	54
		Total	3.58678	.697250	121
	Total	1 Male	3.56707	.710406	164
		2 Female	3.70741	.661819	135
		Total	3.63043	.691300	299

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Timely Feedback scale

F	df1	df2	Sig.
.784	7	291	.602

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + univ + deptB + gender + univ * deptB + univ * gender + deptB * gender + univ *deptB * gender

Tests of Between-Subjects Effects

Dependent Variable:Timely Feedback scale

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	3.652[a]	7	.522	1.094	.367	.026	7.659	.470
Intercept	3440.045	1	3440.045	7214.233	.000	.961	7214.233	1.000
Univ	.027	1	.027	.057	.811	.000	.057	.057
deptB	.497	1	.497	1.043	.308	.004	1.043	.175
Gender	.971	1	.971	2.036	.155	.007	2.036	.296
univ * deptB	.408	1	.408	.856	.356	.003	.856	.152
univ * gender	.032	1	.032	.066	.797	.000	.066	.058
deptB * gender	.614	1	.614	1.289	.257	.004	1.289	.205
univ * deptB * gender	.474	1	.474	.994	.320	.003	.994	.169
Error	138.761	291	.477					
Total	4083.250	299						
Corrected Total	142.413	298						

a. R Squared = .026 (Adjusted R Squared = .002)

b. Computed using alpha = .05

Estimated Marginal Means

1. Institution

Dependent Variable:Timely Feedback scale

Institution	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
2 Uni.B	3.620	.062	3.497	3.742
3 Uni.C	3.640	.058	3.525	3.755

2. Gender

Dependent Variable:Timely Feedback scale

Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1 Male	3.569	.060	3.451	3.687
2 Female	3.691	.061	3.571	3.810

3. Institution * Gender

Dependent Variable:Timely Feedback scale

Institution	Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
2 Uni.B	1 Male	3.548	.092	3.366	3.729
	2 Female	3.692	.084	3.527	3.856
3 Uni.C	1 Male	3.590	.077	3.438	3.742
	2 Female	3.690	.088	3.517	3.863

4. Collapsed Departments

Dependent Variable:Timely Feedback scale

Collapsed Departments	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1 Accounting	3.673	.053	3.569	3.778
2 Other Depts	3.586	.067	3.454	3.718

5. Institution * Collapsed Departments

Dependent Variable:Timely Feedback scale

Institution	Collapsed Departments	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
2 Uni.B	1 Accounting	3.703	.069	3.566	3.839
	2 Other Depts	3.536	.104	3.332	3.740
3 Uni.C	1 Accounting	3.644	.081	3.486	3.803
	2 Other Depts	3.636	.085	3.469	3.803

6. Collapsed Departments * Gender

Dependent Variable:Timely Feedback scale

Collapsed Departments	Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1 Accounting	1 Male	3.564	.073	3.419	3.709
	2 Female	3.783	.077	3.632	3.934
2 Other Depts	1 Male	3.574	.095	3.386	3.761
	2 Female	3.599	.094	3.413	3.784

7. Institution * Collapsed Departments * Gender

Dependent Variable:Timely Feedback scale

Institution	Collapsed Departments	Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
2 Uni.B	1 Accounting	1 Male	3.540	.087	3.368	3.711
		2 Female	3.866	.108	3.654	4.078
	2 Other Depts	1 Male	3.556	.163	3.235	3.876
		2 Female	3.517	.128	3.265	3.770
3 Uni.C	1 Accounting	1 Male	3.588	.118	3.355	3.821
		2 Female	3.700	.109	3.485	3.915
	2 Other Depts	1 Male	3.592	.099	3.398	3.786
		2 Female	3.680	.138	3.408	3.952

Appendix 6.7: Main and Interaction Effects of Institution, Department & Gender Factors on Relevant Feedback Scale

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
Institution	2	Uni.B	151
	3	Uni.C	147
Collapsed Departments	1	Accounting	177
	2	Other Depts	121
Gender	1	Male	164
	2	Female	134

Descriptive Statistics

Dependent Variable:Relevant Feedback scale

Institution	Collapsed Departments	Gender	Mean	Std. Deviation	N
2 Uni.B	1 Accounting	1 Male	3.96032	.551728	63
		2 Female	4.17480	.416585	41
		Total	4.04487	.511587	104
	2 Other Depts	1 Male	3.82407	.589872	18
		2 Female	3.72414	.510361	29
		Total	3.76241	.538095	47
	Total	1 Male	3.93004	.559554	81
		2 Female	3.98810	.506258	70
		Total	3.95695	.534535	151
3 Uni.C	1 Accounting	1 Male	3.82843	.622914	34
		2 Female	3.80342	.487972	39
		Total	3.81507	.551066	73
	2 Other Depts	1 Male	3.77551	.516567	49
		2 Female	3.75333	.678915	25
		Total	3.76802	.571932	74
	Total	1 Male	3.79719	.559500	83
		2 Female	3.78385	.565529	64
		Total	3.79138	.560239	147
Total	1 Accounting	1 Male	3.91409	.577908	97
		2 Female	3.99375	.487137	80
		Total	3.95009	.538764	177
	2 Other Depts	1 Male	3.78856	.533067	67
		2 Female	3.73765	.588679	54
		Total	3.76584	.556768	121
	Total	1 Male	3.86280	.561772	164
		2 Female	3.89055	.543091	134
		Total	3.87528	.552697	298

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Relevant Feedback scale

F	df1	df2	Sig.
1.193	7	290	.307

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + univ + deptB + gender + univ * deptB + univ * gender + deptB * gender + univ * deptB * gender

Tests of Between-Subjects Effects

Dependent Variable:Relevant Feedback scale

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	5.979[a]	7	.854	2.923	.006	.066	20.460	.928
Intercept	3871.028	1	3871.028	13246.501	.000	.979	13246.501	1.000
Univ	1.111	1	1.111	3.803	.052	.013	3.803	.494
deptB	1.937	1	1.937	6.627	.011	.022	6.627	.728
Gender	.018	1	.018	.063	.802	.000	.063	.057
univ * deptB	.953	1	.953	3.260	.072	.011	3.260	.436
univ * gender	.106	1	.106	.364	.547	.001	.364	.092
deptB * gender	.395	1	.395	1.352	.246	.005	1.352	.212
univ * deptB * gender	.410	1	.410	1.401	.237	.005	1.401	.219
Error	84.747	290	.292					
Total	4566.028	298						
Corrected Total	90.726	297						

a. R Squared = .066 (Adjusted R Squared = .043)

b. Computed using alpha = .05

Estimated Marginal Means

1. Institution

Dependent Variable:Relevant Feedback scale

Institution	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
2 Uni.B	3.921	.049	3.825	4.017
3 Uni.C	3.790	.046	3.700	3.881

2. Gender

Dependent Variable:Relevant Feedback scale

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 Male	3.847	.047	3.754	3.940
2 Female	3.864	.048	3.770	3.958

3. Institution * Gender

Dependent Variable:Relevant Feedback scale

Institution	Gender	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
2 Uni.B	1 Male	3.892	.072	3.750	4.034
	2 Female	3.949	.066	3.820	4.079
3 Uni.C	1 Male	3.802	.060	3.683	3.921
	2 Female	3.778	.069	3.642	3.915

4. Collapsed Departments

Dependent Variable:Relevant Feedback scale

Collapsed Departments	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 Accounting	3.942	.042	3.860	4.024
2 Other Depts	3.769	.052	3.666	3.872

5. Institution * Collapsed Departments

Dependent Variable:Relevant Feedback scale

Institution	Collapsed Departments	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
2 Uni.B	1 Accounting	4.068	.054	3.961	4.174
	2 Other Depts	3.774	.081	3.614	3.934
3 Uni.C	1 Accounting	3.816	.063	3.691	3.941
	2 Other Depts	3.764	.066	3.634	3.895

6. Collapsed Departments * Gender

Dependent Variable:Relevant Feedback scale

Collapsed Departments	Gender	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1 Accounting	1 Male	3.894	.058	3.781	4.008
	2 Female	3.989	.060	3.870	4.108
2 Other Depts	1 Male	3.800	.074	3.653	3.946
	2 Female	3.739	.074	3.594	3.884

7. Institution * Collapsed Departments * Gender

Dependent Variable:Relevant Feedback scale

Institution	Collapsed Departments	Gender	Mean	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
2 Uni.B	1 Accounting	1 Male	3.960	.068	3.826	4.094
		2 Female	4.175	.084	4.009	4.341
	2 Other Depts	1 Male	3.824	.127	3.573	4.075
		2 Female	3.724	.100	3.527	3.922
3 Uni.C	1 Accounting	1 Male	3.828	.093	3.646	4.011
		2 Female	3.803	.087	3.633	3.974
	2 Other Depts	1 Male	3.776	.077	3.624	3.928
		2 Female	3.753	.108	3.541	3.966

Appendix 6.8: Main Effect Qualification Factor on Timely & Relevant Feedback Scales

T-Test

Group Statistics

	Collapsed Qualification	N	Mean	Std. Deviation	Std. Error Mean
Timely Feedback scale	1 O'Level	101	3.56931	.624619	.062152
	2 A'Level & Others	198	3.65909	.723916	.051446
Relevant Feedback scale	1 O'Level	100	3.72333	.537703	.053770
	2 A'Level & Others	198	3.94865	.546254	.038821

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Timely Feedback scale	Equal variances assumed	1.649	.200	-1.061	297	.290	-.089784	.084624	-.256323	.076755
	Equal variances not assumed			-1.113	229.331	.267	-.089784	.080682	-.248757	.069189
Relevant Feedback scale	Equal variances assumed	.458	.499	-3.380	296	.001	-.225320	.066666	-.356519	-.094121
	Equal variances not assumed			-3.397	201.580	.001	-.225320	.066320	-.356089	-.094551

Appendix 6.9: Main Effect eTutoring Experience Factor on Timely & Relevant Feedback Scales

T-Test

Group Statistics

	cmpexp6	N	Mean	Std. Deviation	Std. Error Mean
Timely Feedback scale	1 Yes	147	3.66667	.736368	.060735
	2 No	153	3.59150	.644914	.052138
Relevant Feedback scale	1 Yes	147	4.00227	.552766	.045591
	2 No	152	3.75000	.524247	.042522

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Timely Feedback scale	Equal variances assumed	4.071	.045	.942	298	.347	.075163	.079833	-.081944	.232271
	Equal variances not assumed			.939	289.475	.349	.075163	.080044	-.082379	.232706
Relevant Feedback scale	Equal variances assumed	1.659	.199	4.050	297	.000	.252268	.062288	.129686	.374849
	Equal variances not assumed			4.046	294.798	.000	.252268	.062343	.129573	.374962

Appendix 6.10: Main/Interaction Effects of Institution, Department & Gender on Misconception Scale

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
Institution	2	Uni.B	151
	3	Uni.C	148
Collapsed Departments	1	Accounting	178
	2	Other Depts	121
Gender	1	Male	164
	2	Female	135

Descriptive Statistics

Dependent Variable:Misconception scale

Institution	Collapsed Departments	Gender	Mean	Std. Deviation	N
2 Uni.B	1 Accounting	1 Male	3.75132	.677250	63
		2 Female	4.00000	.572519	41
		Total	3.84936	.646757	104
	2 Other Depts	1 Male	3.85185	.596528	18
		2 Female	3.72414	.549668	29
		Total	3.77305	.565115	47
	Total	1 Male	3.77366	.657916	81
		2 Female	3.88571	.575634	70
		Total	3.82561	.621628	151
3 Uni.C	1 Accounting	1 Male	3.50980	.744159	34
		2 Female	3.46667	.573886	40
		Total	3.48649	.653265	74
	2 Other Depts	1 Male	3.59864	.697149	49
		2 Female	3.52000	.844810	25
		Total	3.57207	.745397	74
	Total	1 Male	3.56225	.713645	83
		2 Female	3.48718	.684848	65
		Total	3.52928	.699778	148
Total	1 Accounting	1 Male	3.66667	.707107	97
		2 Female	3.73663	.629630	81
		Total	3.69850	.672001	178
	2 Other Depts	1 Male	3.66667	.676692	67
		2 Female	3.62963	.702397	54
		Total	3.65014	.685635	121
	Total	1 Male	3.66667	.692742	164
		2 Female	3.69383	.659227	135
		Total	3.67893	.676823	299

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Misconception scale

F	df1	df2	Sig.
1.376	7	291	.215

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + univ + deptB + gender + univ * deptB + univ * gender + deptB * gender + univ * deptB * gender

Tests of Between-Subjects Effects

Dependent Variable:Misconception scale

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	8.876[a]	7	1.268	2.891	.006	.065	20.238	.924
Intercept	3531.632	1	3531.632	8051.950	.000	.965	8051.950	1.000
Univ	6.194	1	6.194	14.122	.000	.046	14.122	.963
deptB	.004	1	.004	.010	.920	.000	.010	.051
Gender	.000	1	.000	.000	.998	.000	.000	.050
univ * deptB	.411	1	.411	.938	.334	.003	.938	.162
univ * gender	.240	1	.240	.548	.460	.002	.548	.114
deptB * gender	.692	1	.692	1.578	.210	.005	1.578	.240
univ * deptB * gender	.474	1	.474	1.081	.299	.004	1.081	.179
Error	127.634	291	.439					
Total	4183.333	299						
Corrected Total	136.511	298						

a. R Squared = .065 (Adjusted R Squared = .043)

b. Computed using alpha = .05

Estimated Marginal Means

1. Institution

Dependent Variable:Misconception scale

Institution	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
2 Uni.B	3.832	.060	3.714	3.949
3 Uni.C	3.524	.056	3.413	3.634

2. Gender

Dependent Variable:Misconception scale

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 Male	3.678	.058	3.564	3.791
2 Female	3.678	.058	3.563	3.792

3. Institution * Gender

Dependent Variable:Misconception scale

Institution	Gender	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
2 Uni.B	1 Male	3.802	.088	3.627	3.976
	2 Female	3.862	.080	3.704	4.020
3 Uni.C	1 Male	3.554	.074	3.409	3.700
	2 Female	3.493	.084	3.327	3.659

4. Collapsed Departments

Dependent Variable:Misconception scale

Collapsed Departments	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 Accounting	3.682	.051	3.582	3.782
2 Other Depts	3.674	.064	3.547	3.800

5. Institution * Collapsed Departments

Dependent Variable:Misconception scale

Institution	Collapsed Departments	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
2 Uni.B	1 Accounting	3.876	.066	3.745	4.006
	2 Other Depts	3.788	.099	3.592	3.984
3 Uni.C	1 Accounting	3.488	.077	3.336	3.640
	2 Other Depts	3.559	.081	3.399	3.720

6. Collapsed Departments * Gender

Dependent Variable:Misconception scale

Collapsed Departments	Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1 Accounting	1 Male	3.631	.070	3.492	3.769
	2 Female	3.733	.074	3.588	3.878
2 Other Depts	1 Male	3.725	.091	3.546	3.905
	2 Female	3.622	.090	3.444	3.800

7. Institution * Collapsed Departments * Gender

Dependent Variable:Misconception scale

Institution	Collapsed Departments	Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
2 Uni.B	1 Accounting	1 Male	3.751	.083	3.587	3.916
		2 Female	4.000	.103	3.796	4.204
	2 Other Depts	1 Male	3.852	.156	3.545	4.159
		2 Female	3.724	.123	3.482	3.966
3 Uni.C	1 Accounting	1 Male	3.510	.114	3.286	3.733
		2 Female	3.467	.105	3.261	3.673
	2 Other Depts	1 Male	3.599	.095	3.412	3.785
		2 Female	3.520	.132	3.259	3.781

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
Institution	2	Uni.B	147
	3	Uni.C	144
Gender	1	Male	160
	2	Female	131

Descriptive Statistics

Dependent Variable:Learning Effectiveness scale

Institution	Gender	Mean	Std. Deviation	N
2 Uni.B	1 Male	3.84081	.564585	78
	2 Female	3.96377	.501729	69
	Total	3.89853	.537724	147
3 Uni.C	1 Male	3.92886	.549055	82
	2 Female	3.89919	.532234	62
	Total	3.91609	.540196	144
Total	1 Male	3.88594	.556678	160
	2 Female	3.93321	.515406	131
	Total	3.90722	.538091	291

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Learning Effectiveness scale

F	df1	df2	Sig.
.607	3	287	.611

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + univ + gender + univ * gender

Tests of Between-Subjects Effects

Dependent Variable:Learning Effectiveness scale

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	.607[a]	3	.202	.697	.555	.007	2.090	.197
Intercept	4392.355	1	4392.355	15122.423	.000	.981	15122.423	1.000
univ	.010	1	.010	.034	.854	.000	.034	.054
gender	.156	1	.156	.539	.464	.002	.539	.113
univ * gender	.419	1	.419	1.441	.231	.005	1.441	.223
Error	83.360	287	.290					
Total	4526.472	291						
Corrected Total	83.967	290						

a. R Squared = .007 (Adjusted R Squared = -.003)

b. Computed using alpha = .05

Estimated Marginal Means

1. Institution

Estimates

Dependent Variable:Learning Effectiveness scale

Institution	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
2 Uni.B	3.902	.045	3.815	3.990
3 Uni.C	3.914	.045	3.825	4.003

Pairwise Comparisons

Dependent Variable:Learning Effectiveness scale

(I) Institution	(J) Institution	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
2 Uni.B	3 Uni.C	-.012	.064	.854	-.137	.113
3 Uni.C	2 Uni.B	.012	.064	.854	-.113	.137

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Univariate Tests

Dependent Variable:Learning Effectiveness scale

	Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	.010	1	.010	.034	.854	.000	.034	.054
Error	83.360	287	.290					

The F tests the effect of Institution. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

2. Gender

Estimates

Dependent Variable:Learning Effectiveness scale

Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1 Male	3.885	.043	3.801	3.969
2 Female	3.931	.047	3.839	4.024

Pairwise Comparisons

Dependent Variable:Learning Effectiveness scale

(I) Gender	(J) Gender	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	95% Confidence Interval for Difference[a] Upper Bound
1 Male	2 Female	-.047	.064	.464	-.172	.078
2 Female	1 Male	.047	.064	.464	-.078	.172

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Univariate Tests

Dependent Variable:Learning Effectiveness scale

	Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	.156	1	.156	.539	.464	.002	.539	.113
Error	83.360	287	.290					

The F tests the effect of Gender. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

3. Institution * Gender

Dependent Variable:Learning Effectiveness scale

Institution	Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
2 Uni.B	1 Male	3.841	.061	3.721	3.961
	2 Female	3.964	.065	3.836	4.091

3 Uni.C	1 Male	3.929	.060	3.812	4.046
	2 Female	3.899	.068	3.764	4.034

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
Institution	2	Uni.B	148
	3	Uni.C	147
Collapsed Departments	1	Accounting	176
	2	Other Depts	119
Gender	1	Male	161
	2	Female	134

Descriptive Statistics

Dependent Variable:Cognitive Visibility & Learning scale

Institution	Collapsed Departments	Gender	Mean	Std. Deviation	N
2 Uni.B	1 Accounting	1 Male	4.04839	.521895	62
		2 Female	4.12000	.531664	40
		Total	4.07647	.524296	102
	2 Other Depts	1 Male	3.67059	.628256	17
		2 Female	3.93103	.573899	29
		Total	3.83478	.601190	46
	Total	1 Male	3.96709	.564261	79
		2 Female	4.04058	.553683	69
		Total	4.00135	.558665	148
3 Uni.C	1 Accounting	1 Male	4.24706	.703341	34
		2 Female	4.19500	.462961	40
		Total	4.21892	.582078	74
	2 Other Depts	1 Male	3.97500	.727134	48
		2 Female	3.97600	.798582	25
		Total	3.97534	.746804	73
	Total	1 Male	4.08780	.725614	82
		2 Female	4.11077	.617486	65
		Total	4.09796	.677775	147
Total	1 Accounting	1 Male	4.11875	.596536	96
		2 Female	4.15750	.496768	80
		Total	4.13636	.552228	176
	2 Other Depts	1 Male	3.89538	.710728	65
		2 Female	3.95185	.680660	54
		Total	3.92101	.694881	119
	Total	1 Male	4.02857	.652249	161
		2 Female	4.07463	.584343	134
		Total	4.04949	.621709	295

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Cognitive Visibility & Learning scale

F	df1	df2	Sig.
1.479	7	287	.174

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + univ + deptB + gender + univ * deptB + univ * gender + deptB * gender + univ * deptB * gender

Tests of Between-Subjects Effects

Dependent Variable:Cognitive Visibility & Learning scale

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	5.622[a]	7	.803	2.134	.040	.049	14.938	.807
Intercept	4143.148	1	4143.148	11008.452	.000	.975	11008.452	1.000
univ	1.555	1	1.555	4.131	.043	.014	4.131	.526
deptB	4.482	1	4.482	11.908	.001	.040	11.908	.930
gender	.316	1	.316	.840	.360	.003	.840	.150
univ * deptB	.023	1	.023	.061	.805	.000	.061	.057
univ * gender	.588	1	.588	1.562	.212	.005	1.562	.238
deptB * gender	.234	1	.234	.623	.431	.002	.623	.123
univ * deptB * gender	.074	1	.074	.196	.658	.001	.196	.073
Error	108.016	287	.376					
Total	4951.160	295						
Corrected Total	113.637	294						

a. R Squared = .049 (Adjusted R Squared = .026)

b. Computed using alpha = .05

Estimated Marginal Means
1. Institution

Estimates

Dependent Variable:Cognitive Visibility & Learning scale

Institution	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
2 Uni.B	3.943	.056	3.832	4.053
3 Uni.C	4.098	.052	3.996	4.201

Pairwise Comparisons

Dependent Variable:Cognitive Visibility & Learning scale

(I) Institution	(J) Institution	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
2 Uni.B	3 Uni.C	-.156*	.077	.043	-.307	-.005
3 Uni.C	2 Uni.B	.156*	.077	.043	.005	.307

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Univariate Tests

Dependent Variable:Cognitive Visibility & Learning scale

	Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	1.555	1	1.555	4.131	.043	.014	4.131	.526
Error	108.016	287	.376					

The F tests the effect of Institution. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

2. Collapsed Departments

Estimates

Dependent Variable:Cognitive Visibility & Learning scale

Collapsed Departments	Mean	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
1 Accounting	4.153	.047	4.059	4.246
2 Other Depts	3.888	.060	3.770	4.007

Pairwise Comparisons

Dependent Variable:Cognitive Visibility & Learning scale

(I) Collapsed Departments	(J) Collapsed Departments	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1 Accounting	2 Other Depts	.264*	.077	.001	.114	.415
2 Other Depts	1 Accounting	-.264*	.077	.001	-.415	-.114

Based on estimated marginal means

Estimates

Dependent Variable:Cognitive Visibility & Learning scale

Collapsed Departments	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1 Accounting	4.153	.047	4.059	4.246

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Univariate Tests

Dependent Variable:Cognitive Visibility & Learning scale

	Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	4.482	1	4.482	11.908	.001	.040	11.908	.930
Error	108.016	287	.376					

The F tests the effect of Collapsed Departments. This test is based on the linearly independent pairwise comparisons among the estimated marginal means. a. Computed using alpha = .05

3. Gender

Estimates

Dependent Variable:Cognitive Visibility & Learning scale

Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1 Male	3.985	.054	3.878	4.092
2 Female	4.056	.054	3.949	4.162

Pairwise Comparisons

Dependent Variable:Cognitive Visibility & Learning scale

(I) Gender	(J) Gender	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	95% Confidence Interval for Difference[a] Upper Bound
1 Male	2 Female	-.070	.077	.360	-.221	.081
2 Female	1 Male	.070	.077	.360	-.081	.221

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Univariate Tests

Dependent Variable:Cognitive Visibility & Learning scale

	Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	.316	1	.316	.840	.360	.003	.840	.150
Error	108.016	287	.376					

The F tests the effect of Gender. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

4. Institution * Collapsed Departments

Dependent Variable:Cognitive Visibility & Learning scale

Institution	Collapsed Departments	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
2 Uni.B	1 Accounting	4.084	.062	3.962	4.207
	2 Other Depts	3.801	.094	3.616	3.985
3 Uni.C	1 Accounting	4.221	.072	4.080	4.362
	2 Other Depts	3.976	.076	3.827	4.124

5. Institution * Gender

Dependent Variable:Cognitive Visibility & Learning scale

Institution	Gender	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
2 Uni.B	1 Male	3.859	.084	3.694	4.025
	2 Female	4.026	.075	3.878	4.173
3 Uni.C	1 Male	4.111	.069	3.976	4.246
	2 Female	4.085	.078	3.932	4.239

6. Collapsed Departments * Gender

Dependent Variable:Cognitive Visibility & Learning scale

Collapsed Departments	Gender	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1 Accounting	1 Male	4.148	.065	4.019	4.277
	2 Female	4.157	.069	4.022	4.293
2 Other Depts	1 Male	3.823	.087	3.652	3.993
	2 Female	3.954	.084	3.789	4.118

7. Institution * Collapsed Departments * Gender

Dependent Variable:Cognitive Visibility & Learning scale

Institution	Collapsed Departments	Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
2 Uni.B	1 Accounting	1 Male	4.048	.078	3.895	4.202
		2 Female	4.120	.097	3.929	4.311
	2 Other Depts	1 Male	3.671	.149	3.378	3.963
		2 Female	3.931	.114	3.707	4.155
3 Uni.C	1 Accounting	1 Male	4.247	.105	4.040	4.454
		2 Female	4.195	.097	4.004	4.386
	2 Other Depts	1 Male	3.975	.089	3.801	4.149
		2 Female	3.976	.123	3.735	4.217