UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
School of Electronics and Computer Science

A competency model for semi-automatic question generation in adaptive assessment

by

Onjira Sitthisak

Thesis for the degree of Doctor of Philosophy

May 2009

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING

SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

A COMPETENCY MODEL FOR SEMI-AUTOMATIC QUESTION GENERATION IN ADAPTIVE ASSESSMENT

by Onjira Sitthisak

The concept of competency is increasingly important since it conceptualises intended learning outcomes within the process of acquiring and updating knowledge. A competency model is critical to successfully managing assessment and achieving the goals of resource sharing, collaboration, and automation to support learning. Existing e-learning competency standards such as the IMS Reusable Definition of Competency or Educational Objective (IMS RDCEO) specification and the HR-XML standard are not able to accommodate complicated competencies, link competencies adequately, support comparisons of competency data between different communities, or support tracking of the knowledge state of the learner.

Recently, the main goal of assessment has shifted away from content-based evaluation to intended learning outcome-based evaluation. As a result, through assessment, the main focus of assessment goals has shifted towards the identification of learned capability instead of learned content. This change is associated with changes in the method of assessment.

This thesis presents a system to demonstrate adaptive assessment and automatic generation of questions from a competency model, based on a sound pedagogical and technological approach.

The system's design and implementation involves an ontological database that represents the intended learning outcome to be assessed across a number of dimensions, including level of cognitive ability and subject matter content. The system generates a list of the questions and tests that are possible from a given learning outcome, which may then be used to test for understanding, and so could determine the degree to which learners actually acquire the desired knowledge.

Experiments were carried out to demonstrate and evaluate the generation of assessments, the sequencing of generated assessments from a competency data model, and to compare a variety of adaptive sequences. For each experiment, methods and experimental results are described. The way in which the system has been designed and evaluated is discussed, along with its educational benefits.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Onjira Sitthisak, declare that the thesis entitled "A competency model for semi-automatic question generation in adaptive assessment" and work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as :

1. Gilbert, L., Sitthisak, O., Sim, Y. W., Wang, C., and Wills, G. (2006) From collaborative virtual research environment to teaching and learning. In TENCompetence Workshop: Learning Networks for Lifelong Competence Development, Sofia, Bulgaria.
2. Sitthisak, O., Gilbert, L. and Davis, H. C. (2007) Towards a competency model for adaptive assessment to support lifelong learning. In: *TENCompetence Workshop on Service Oriented Approaches and Lifelong Competence Development Infrastructures*, 11-12 January 2007, Manchester, UK.

3. Sitthisak, O., Gilbert, L., Zalfan, M. T. and Davis, H. C. (2007) Interactivity within IMS learning design and question and test interoperability. In: *The 3rd International Conference on Web Information Systems and Technologies (WEBIST)*, March 3-6, Barcelona, Spain.
4. Gilbert, L. and Sitthisak, O. (2007) Pedagogically informed metadata content and structure for learning and teaching. In: *TENCompetence Open Workshop on Current research on IMS Learning Design and Lifelong Competence Development Infrastructures*, 21-22 June 2007, Barcelona, Spain.
5. Sitthisak, O., Gilbert, L., Davis, H. C. and Gobbi, M. (2007) Adapting health care competencies to a formal competency model. In: *The 7th IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, July 18-20, 2007, Niigata, Japan.
6. Sitthisak, O., Gilbert, L. and Davis, H. C. (2007) Transforming a competency model to assessment items. In: *PROLIX Workshop 2007 in conjunction with EC-TEL07*, 18 September 2007, Crete, Greece.
7. Sitthisak, O., Gilbert, L. and Davis, H. (2008) Transforming a competency model to assessment items. In: *4th International Conference on Web Information Systems and Technologies (WEBIST)*, May 4-7, 2008, Funchal, Madeira - Portugal.
8. Sitthisak, O., Gilbert, L. and Davis, H. (2008) Deriving e-assessment from a competency model. In: *The 8th IEEE International Conference on Advanced Learning Technologies (ICALT 2008)*, July 1-5, 2008, Santander, Cantabria, Spain.
9. Sitthisak, O., Gilbert, L. and Davis, H. (2008) An evaluation of pedagogically informed parameterised questions for self assessment. *Learning, Media and Technology*, vol.33, 3, pp. 235-248.
10. Sitthisak, O., Gilbert, L., and Davis, H. C. (2009) *Transforming a Competency Model to Parameterised Questions in Assessment*. vol. 18, Lecture Notes in Business Information Processing: Springer-Verlag Berlin Heidelberg.

Signed:

Date: 13 May 2009

Acknowledgements

During the period of my PhD, I have had cause to be grateful to many people for their support. In particular, I would like to thank the following people.

My parents and *Dr. Damrongsak Tongsomporn*, for always being there to love me and for providing mental support throughout my lifetime;

My supervisors, *Dr. Hugh Davis* and *Lester Gilbert*, for being my supporters over the period of my PhD, for giving me their wonderful supervision and their time to read my work, for teaching me the importance of being systematic, how to prove something, and how to write like a scientist;

My colleagues in Learning Societies Lab, *Joseph Price* and *Dr. Hend S. Al-Khalida* for proofreading my work in its early stages, *Reena Pau* for giving me a rich experience at the University of Southampton and helping me improve my English, and other staff and colleagues for being good users of my system;

The Royal Thai Government, for full funding during my PhD life;

Quintin Gee, for proofreading my thesis.

Special thanks to *Lester Gilbert*, for being an excellent teacher, for giving assignments as a demonstrator and project manager so that I gained valuable working experience, and for giving me spiritual and intellectual guidance and challenging me to think in new ways.

Definitions and Abbreviations Used

RDCEO	Reusable Competency Definitions Information Model
SCORM	Sharable Content Object Reference Model
RDF	Resource Description Framework
OWL	Web Ontology Language
SKOS	Simple Knowledge Organisation System
ASDEL	Assessment Delivery Engine for QTIv2 questions

Intended learning outcome, learning outcome, competence and competency are terms used throughout this thesis to mean similar things. In relevant sections, they are more exactly defined and disambiguated as necessary.

Chapter 1 Introduction

1.1 Overview of research

Recently, emphasis has shifted away from content-based education, which describes what instructors do and the content of material presented during classroom instruction, to intended learning outcomes, which describe what learners can do as a result of their educational experiences. This change is associated with changes in the main goal of assessment.

Assessment is part of the developmental process of learning (Kommers, Grabinger and Dunlap, 1996) and is related to the accomplishment of learning outcomes. Through assessment, learners are able to identify what they have already learned and which are their strengths and weaknesses, to observe their personal learning progress, and to decide how to further direct their learning process. This involves a learner's competency.

The question arises "Can we create assessments from statements of competency?" In order to answer this question, a machine-processable competency statement is crucial. Of course, this competency statement should be understood by humans as well. If the competency statement is machine-readable, the question "Can assessment systems automatically convert this statement into questions?" is significant. If assessments can be automatically generated, the question "Can these be adaptive to the current knowledge level and goals of the learner?" leads to a critical research question.

In this study, machine-processable competency statements are supported by ontologies and taxonomies of competence. Machine processing can offer interoperable, portable, reusable resources and applications that are pedagogically effective for e-learning and assessment.

A competency statement which can be read, processed, and interpreted by machine contributes to the automatic generation of questions and sequences of questions, and offers a semantic structure for further processing. The use of a competency model expresses the achieved competencies of learners, and provides adaptation of the system proposed here with adaptive assessments.

1.2 Research objectives

The objectives of this research are to:

- Develop a model of competency that supports the adaptation and automatic generation of questions.
- Investigate what information should be included in the model to represent a learner's competency, and how this information should be represented in order to be machine-readable and understood by humans.
- Demonstrate a design and implementation of the model, to represent the intended learning outcomes to be assessed across a number of dimensions.
- Demonstrate an example of an adaptive assessment system using the competency model for automatically generating questions.

1.3 Thesis structure

This thesis is divided into ten chapters as follows:

Chapter 2 reviews the major concepts of learning theory. Using different learning theories for different purposes is discussed. E-learning components, instructional designs, and the development of effective instructional process are next briefly outlined. The next topic is the instructional content and ability matrix, where many methods of representation of this matrix, and benefits of using the matrix, are discussed.

Chapter 3 introduces the concept of effective questions and discusses the application of automatic question generating systems. The benefits of adaptive hypertext and adaptive assessment are considered. This is followed by adaptation techniques and the application of adaptive assessment systems. Finally, some of the problems faced by the authors of adaptive assessment applications are highlighted.

Chapter 4 presents the definition of knowledge and describes the processes of knowledge engineering, and the processes involved in the acquisition and representation of knowledge. A detailed description, the basic technologies, and architecture of the semantic web are given. Then follows the definition and construction of ontologies. The chapter ends with issues arising from applying ontologies in education and e-learning.

Chapter 5 introduces competency and e-learning models. The chapter begins by establishing some understanding about what is meant by competency. Relevant competency and e-learning standards are described.

Chapter 6 describes the development of the proposed competency model. The criteria for a good competency model are discussed, and the requirements for competency standards are detailed. These standards are compared. This chapter then introduces an improved competency model produced by comparing the competency standards against the desired taxonomy of competence. Next is a comparison of usability criteria within the existing competency standards and comparing the standards with the improved competency model. The criteria used for comparison are reusability, interoperability, equivalency and similarity, measurability and measurable behaviours, defining domain and scope of ontology, and personalisation.

Chapter 7 describes the process of designing and developing the COMBA system. The processes comprise data creation, representation and storage, methods of generating, and methods of question delivery. The architecture of the system is presented. Expressing the basic structure of a domain using ontologies is described. This chapter also includes a description of storing and querying OWL ontologies. Methods of generating questions from given question templates and adaptive sequence are explored. Finally, a stand-alone web application, ASDEL, is described, which is responsible for delivering the tests to the learners.

Chapter 8 evaluates the feasibility of the proposed model and architecture for generating questions, and to deliver adaptive questions to learners. Three experiments have been performed to demonstrate the generation of assessments, to demonstrate the sequencing of generated assessments from a competency data model, and to compare a variety of adaptive sequences. The chapter ends with some justification of the experiments' results.

Chapter 9 evaluates the output of the research prototype and discusses these results. The quality of the questions generated is considered, the usefulness and adaptiveness of the COMBA generated test discussed. The opinion of students and experts about the adaptive sequencing of the COMBA generated test is examined, as is the relationship between a variety of adaptive sequences and their pedagogical effectiveness.

Chapter 10 discusses issues arising from the development of the COMBA system. The chapter ends with recommendations for future work emanating from this thesis.

Chapter 2 Instructional Theory

2.1 Introduction

There is a growing awareness of the need for effective dialogue among learners, teachers, educational researchers, and developers of standards and systems, to ensure that e-learning systems are usable by learners and teachers in flexible and appropriate ways, and to ensure that teachers can design activities that meet the needs of learners. The aim of this chapter is to review, within the context of current learning theories, the concepts of e-Learning components, instructional design, and instructional content and ability matrix.

2.2 Learning theories

In the paper which described the freedom to learn, Rogers (1983) noted that learning is concerned with "what I am interested in, what I need and what I want to know". Learning theories are an attempt at describing how learners learn. With a learning theory as a foundation, instruction can be structured around making learning most effective (Ertmer and Newby, 1993). Current theories of learning include behaviourism, cognitivism, and constructivism (Mergel, 1998; Ryder, 2006). For example, a behaviourist approach might assess learners to determine a starting point for instruction, while a cognitivist approach might research the learners to determine their predisposition to learning. A constructivist approach produces a product that is much more facilitative in nature than prescriptive. The workflow of content is determined by the learner, and assessment is much more subjective because it does not depend on specific quantitative criteria, but rather the process and self-evaluation of the learner.

Recently, there have been many e-learning models and frameworks which claim to employ a constructivist approach (Mayes and Freitas, 2004). Furthermore, in the study of a research-led approach developed by Davis and White (2001), they presented a teaching innovation, motivated from a social constructivist perspective that moved away from a world of passive learning into a space where the learners learn by constructing knowledge and understanding by interaction with others.

However, an instructional designer can use different learning theories for different purposes based on particular objectives and models.

2.3 e-Learning components

There are many published papers on e-learning components, such as Kolås and Staupe (2004) and JISC (2004), because the authors considered that an e-learning system should focus on more than technologies. In addition, the Joint SFEFC/SHEFC e-learning Group (Joint SFEFC/SHEFC E-Learning Group, 2003) noted that e-learning is fundamentally about learning, not about technology. Therefore, strategic development of e-learning should be based on the needs and demands of learners and the quality of their educational experience. As a result, many researchers have attempted to discover e-learning factors that influence instructional design. A summary of e-learning components will be covered in the following section.

Kolås and Staupe (2004) presented e-learning components with four different aspects: media, content, administration, and methods – emphasising pedagogical methods. Their idea is that teachers are used to finding the best methods of teaching their subject, and likewise online teachers should have the same opportunity. Hence, the main focus for the e-learning component is the pedagogy, not the technology.

JISC (2004) focused mainly on learning activities. A learning activity can be defined as an interaction between a learner and an environment, leading to a planned outcome. As a result, e-learning components consist of a learner, the learning environment, the intended learning outcomes, the learning activities, and the approach taken that relates to the learner's needs. Further, Conole and Fill (2005) defined the notion of learning activity as composed of a context, a learning and teaching approach, and a task. The essence of the learning activity is that it must have one or more learning outcomes associated with it.

2.4 Instructional design

According to Smith and Ragan (1999), instruction is a part of education because all instruction consists of experiences leading to learning. Thus, effective and efficient instruction has to be designed accurately and reliably to achieve the intended learning outcomes. Many instructional design theories have been formulated in order to guide the process of learning and support the notion that an essential element in instruction is the application of learning theories (Baruque and Melo, 2003). In a paper which described a new paradigm of instructional theories, Reigeluth (1999) stated that instructional design theories were design-oriented. Methods of instruction should be used and broken into simple component methods.

Berger and Kam (1996) defined instructional design as process, discipline, science and reality. It is concerned with the development process and instructional content towards identified learning goals using learning and instructional theories to ensure the efficiency and effectiveness of instruction. The instructional design is mentioned in the following two topics: the instructional process, and learning theories.

2.5 Developing effective instructional processes

The development of an effective instructional process is presented in terms of the instructional design model. These are more commonly known eponymously such as the Dick and Carey Model (1990), the Pebble-in-the-Pond Design model (Merrill, 2002), etc. They may also be presented in the form of a development methodology such as the e-learning system engineering methodology (Gilbert, Sim and Wang, 2005b), the E-Learning Engineering methodology (2004), the ADDIE model (Dick and Cary, 1996), and the Rapid eLearning Development Process (Prakash *et al.*, 2000; XIA Systems Corporations, 2005).

Activities in these models and methodologies are similar to the principles of software engineering which technical developers know in Structured Systems Analysis and Design Methodology (SSADM) (Weaver, 1993). However, the instructional developers have to modify the whole process of the software engineering approach to different instructional design theories depending on the view of each theory of the context of use, instructional objective, and point of relevance.

The instructional model and framework Nine Steps of Instruction (Gagne, 1985), Laurillard's conversational framework (Laurillard, 2002), and a model of learning transaction (Gilbert, Sim and Wang, 2005a), agree that a good instructional model or framework helps teachers to structure the material, and the programmer to develop pedagogically suitable units of learning. It also assists the instructional designer to design appropriate interactions and to provide useful toolkit support.

To summarise, an instructional design model gives a structure and meaning to an instructional design problem, enabling the instructional designers to discuss their design task with an appearance of understanding. Models help them visualise the problem, breaking it down into discrete and manageable units.

2.6 Instructional content and learner ability

The issues of instructional content and learning outcomes as parts of instructional design may be presented in the form of a matrix of content by ability (Kemp, Morrison and Ross, 1998; Krathwohl and Anderson, 2002; Merrill, 1999). There are many methods of representation such as the combination of Merrill's analysis of subject content and analysis of cognitive capability, illustrated in Table 2-1 (Gilbert and Gale, 2007). Figure 2-1 shows the type of intended outcomes or objective in well-known models.

Subject Content	Cognitive capability			
	Remember instance	Remember generality	Use	Find
Fact	Fact pair Name			
Concept	Example Name	Define Example State definition	Define Example Classify	Explore categories Invent definition
Procedure	Demonstration Rehearse	Definition Demonstration State steps	Definition Demonstrations Demonstrate	Explore procedures Devise procedure
Principle	Explanation Explain	Definition Explanation State cause-effect relationships	Definition Explanations Predict	Explore problems Discover principle

Table 2-1 Form of content by ability matrix

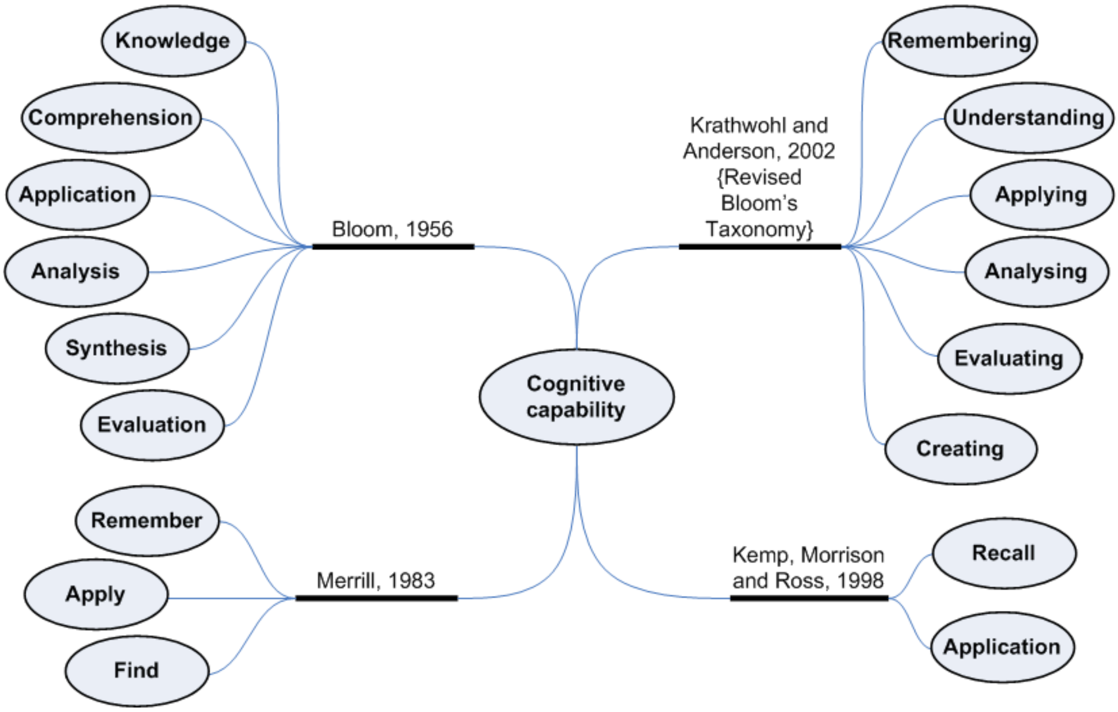


Figure 2-1 Taxonomies of cognitive capability in well-known models

The classification of content and ability come together because of the design requirement to align subject content and required ability with assessment techniques. Teachers may also use the matrix to examine current learning outcomes in units of learning and teaching, and to revise the intended learning outcomes so that they will align with one another, and with the unit's assessments. Further, using the matrix may give teachers a place to start when revising units to better align with new standards-based requirements.

2.7 Summary

An instructional design model gives a structure and meaning to an instructional design problem, enabling the instructional designers to discuss their design task with an appearance of understanding. Models help them visualise the problem, breaking it down into discrete and manageable units.

Assessment is a crucial component of learning. Learners can learn by asking themselves questions and attempting to answer them. However, creating effective questions is time-consuming because it may require considerable resources and the skill of critical thinking. Questions need careful construction to accurately represent the intended learning outcomes and the subject matter involved. There are many systems currently available which generate questions automatically, and these are confined to specific domains. The following chapter presents the criteria for effective questions, the state of the art of automatic question generating systems, and the applications of assessment systems.

Chapter 3 Assessment Systems

3.1 Introduction

Before starting to explore the application of assessment systems, the concept of effective questions and the application of automatic question generating systems are introduced and discussed. This is followed by consideration of the benefits of adaptive assessment. Finally, some problems of adaptive assessment applications are described.

3.2 Effective questions

Questioning is useful because it challenges learners to respond and it reveals learners' abilities to reason, create, analyse, synthesise, and evaluate. Before starting to write questions, the criteria for effective questions should be considered. The following criteria are summarized from Horton (2006), Silyn-Roberts (1996), and McComas and Abraham (2005).

In this research, the use of a multiple choice question was focused. A question should relate to the learning outcomes being measured. Hence, this research involves a method of transforming intended learning outcomes to questions. Question phrasing should be precise, clear, and easy to understand by using the simplest possible language. Good questions should appropriately challenge learners in order to stimulate them to think more deeply about the subject matter. Finally, a good question should help the learner to identify where further study may be useful.

3.3 Automatically generating questions

There are currently many systems available to generate questions automatically; these are however confined to specific domains. A number of pioneering systems such as Problets (Dancik and Kumar, 2003), ILE (Cristea and Tuduce, 2005), QuizPACK (Brusilovsky and Sosnovsky, 2005), and Jeliot 3 (Myller, 2007), explored the use of automatic generation of questions using parameterised templates. The basic concept uses templates instantiated with random values to generate the questions. A question's template is able to produce a large number of different questions.

Problets and Jeliot 3 generate questions about programming using computer language templates. The question generation of Problets is language independent, whereas Jeliot currently supports only Java. Problets and Jeliot are self-contained, lacking interoperability with other systems such as institutional-wide e-learning systems.

ILE is a tool that automatically generates exercises for the special case of electric AC circuit problems, given global parameters such as the number of nodes and number of branches.

QuizPACK works on automatic evaluation of code-execution questions. A teacher provides the core content of a question, a parameterised fragment of code to be executed, and a variable within that code. QuizPACK randomly generates the value of the question parameter, creates a presentation of the resulting question, and runs the presented code in order to generate the correct answer.

These applications of parameterised questions were developed for computer programming and other math-related subjects. A correct answer to a parameterised question can be calculated by a formula or executed by a standard language complier without the need for a teacher or author to provide it. Currently, such systems offer remarkable automatic generation of questions, but only for specific domains, and lack integration, interoperability, portability and reusability.

3.4 Adaptive hypertext

In the paper which examined the need for an open hypertext protocol, Reich and Davis (1999) stated that the limitations of hypertext systems were that they were monolithic and closed. At that time, the Multimedia Research group at the University of Southampton produced Microcosm, a prototypical open hypertext system, to characterize openness and interoperability (Davis, 1999). As a result, Microcosm was the first of a new generation of hypertext authoring environments. Research at University of Southampton has led to the development of the Southampton Framework for Agent Research (SoFAR) (Moreau *et al.*, 2000). Using this platform, an implementation-oriented framework for producing adaptive hypermedia system has been developed as part of the Agent-Based framework for Adaptive Hypermedia (ABAH). This framework offers flexible solutions to developing adaptive hypermedia and adaptive web-based applications and can be used in conjunction with existing models of adaptive hypermedia (Bailey, 2001). At that time, many research fields dealt with hypertext, such as adaptive assessment and adaptive hypertext, which will be described in the following section.

The need to overcome problems with non-adaptive hypertext systems has motivated the evolution of adaptive hypertext and hypermedia research (Brusilovsky, 2001). The goal of adaptive hypermedia is to improve the usability of hypermedia through the automatic adaptation of hypermedia applications to individual users (De Bra, 2000). Many adaptive hypermedia systems have been developed based on a detailed taxonomy of methods and techniques developed by Brusilovsky (2001a) such as ELM-ART (Weber and Brusilovsky, 2001), INSPIRE (Papanikolaou *et al.*, 2002) and AHA! (De Bra *et al.*, 2003).

In an adaptive hypertext system, adaptation can reduce a cognitive overload problem but it also can induce loss in hyperspace in the following cases. Firstly, document grouping in the classification of hypertext can occur. In the case that the document has its context relating to many subjects, the classification tools may put this document in all subjects or in a few main subjects. Secondly, wrong initial user data for adaptation can possibly appear. If a system of data elicitation is incompletely designed, it will provide wrong user information to the data analysis and affect the adaptation. Thirdly, if the user's mental model is created incorrectly because of implementation techniques, it will provide wrong information to the adaptation. Fourthly, if some useful information was possibly filtered out, the user will have no chance to see and make a decision whether he/she is interested in the information or not. Therefore, these cases should be considered when an adaptive hypertext system is developed.

3.5 Adaptation techniques

There are two adaptation techniques; presentational adaptation and navigational adaptation (Brusilovsky, 1996) as illustrated in Figure 3-1.

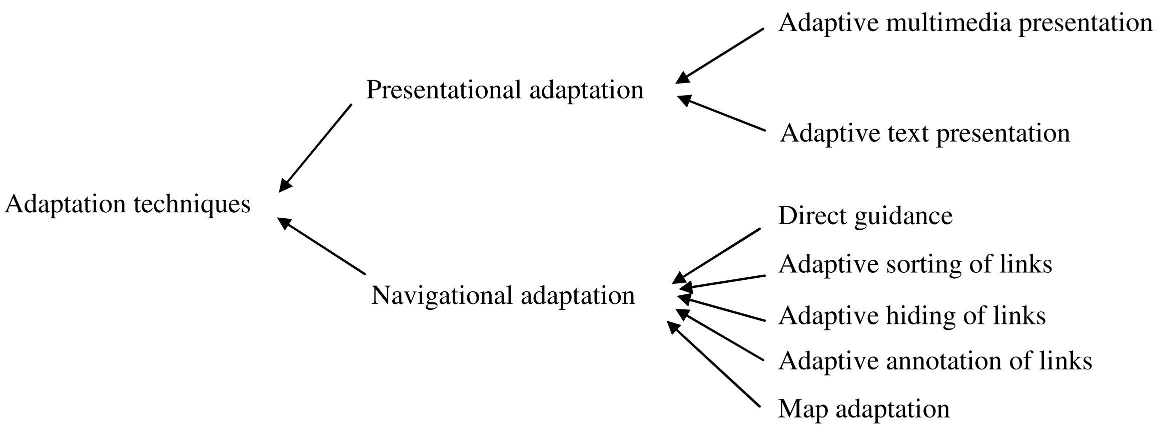


Figure 3-1 Adaptation technologies (Brusilovsky, 1996)

3.5.1 Presentational adaptation

The idea of various adaptive presentation techniques is to adapt the content of a page accessed by a particular user to the current knowledge level, goals, and other characteristics of the user. For example, a qualified user can be provided with more detailed and deep information, while a novice can receive additional explanation. There are many works based on adaptive text such as Personal-reader (Cheniti-Belcadhi and Braham, 2004) and INSPIRE (Papanikolaou *et al.*, 2002).

3.5.2 Navigational adaptation

The idea of adaptive navigation support techniques is to help users to find their paths in hyperspace by adapting the style of link presentation to the goals, knowledge, and other characteristics of an individual user. Adaptive navigation support techniques can be classified into several groups, according to the method they use to adapt presentation of links. These groups of techniques are considered as different technologies for adapting link presentation. The most popular technologies are direct guidance, sorting, hiding, and annotation such as ISIS-Tutor (Brusilovsky and Pesin, 1994), Personal-reader (Cheniti-Belcadhi and Braham, 2004), and INSPIRE

(Papanikolaou *et al.*, 2002). This technique was implemented in this study by adapting assessment paths according to a learner's previous answer.

3.6 Adaptive assessment and its applications

There are a number of adaptive assessment methods and technologies that can be used to assess learners' strengths and weaknesses based on item-by-item and learner responses. These allow learners to be tested on materials at their level. Adaptive assessments change their behaviour and structure depending on the learner's responses and detected abilities.

The key idea of an adaptive assessment system is that questions are selected by the computer to individually match the ability level of each learner. In this approach, the test is tailored to each learner (Askins, 2004; Way, 2005). Adaptive assessment aims to assess a learner's competency by posing a minimum number of questions in order to decrease test length, which is one of the main goals in adaptive assessment (Welch and Frick, 1993). Another main goal includes offering personalized support according to the needs and ability of each learner (Brusilovsky, 1996). The system may skip over what learners have learned and find out what they should learn further. As a result, most existing test engines present questions according to the level of the learner's abilities in order to eliminate too easy or too difficult questions (Askins, 2004; Gouli *et al.*, 2004). Therefore, adaptive questioning is an efficient and effective mean of knowledge-based assessment.

Work related to the proposed approach can be found in the areas of adaptive assessment system. Many adaptive assessment systems have been developed such as Personal-reader (Cheniti-Belcadhi, Henze and Braham, 2005), INSPIRE (Papanikolaou and Grigoriadou, 2003), COMPASS (Gouli *et al.*, 2004), SIETTE (Conejo *et al.*, 2004), AthenaQTI (Tzanavari, Retalis and Pastellis, 2004), and CosyQTI (Lalos, Retalis and Psaromiligkos, 2005). These systems are described below.

1. Personal-reader

This system is developed to personalize a learner's assessment at each moment of the learning process. In addition, this system provides the functionality for selecting the learning goal of each learner in order to personalize learning content and assessment content. The learning content and assessment content are adapted according to the current learner performance, and the current learning content presented to the learner. There are two types of learning content: atomic learning object and linear learning object. In the case where the learner gives wrong answers, the assessment framework should detect the atomic learning objects that have to be studied again, highlights them and gives, if necessary, some additional links that could be used to better understand the current lesson. In the case where the answers are correct, the learner is allowed to continue. Then new course material is generated in the next linear learning object. This system uses the TRIPLE language and First Order Logic in order to define the adaptive rule. This work focuses on interoperability and reusability of the assessment information among different systems by using IEEE LOM, IMS QTI, IEEE PAPI and Semantic Web technologies. In summary, this system still has problems of representing learning knowledge and has difficulty with problem solving.

2. INSPIRE (INtelligent System for Personalized Instruction in a Remote Environment)

Based on the learning goal that the learner selects, INSPIRE generates lessons that correspond to specific learning outcomes, accommodating the learner's knowledge level and learning style. Thus, aiming at individualizing instruction, the system generates lesson plans tailored to the needs, preferences and knowledge level of each individual learner by making use of information about the learner gathered through their interaction. The assessment mechanism poses questions by considering the difficulty level of each question and the current learner's proficiency. The level of performance is represented by Bloom's taxonomy (Bloom and Krathwohl, 1956). Then it provides meaningful feedback. The wrong answer is given with hints and links to locations of supplementary knowledge. Hence, the learner is offered a self-assessment test and a summative assessment test. This system focuses on techniques of selecting tests by using Item Response Theory. In summary, this system still has the problems of collaboration with many teachers, and the use of numerous parameters associated with each question for teachers who are usually practically focused and who would have difficulty with controlling user interaction.

3. COMPASS (Concept MaP ASSessment tool)

COMPASS is an adaptive web-based concept map assessment tool. Based on an assessment goal that the learner selects from a set of proposed goals, COMPASS engages learners to the assessment and learning process through a set of assessment activities. The activities address specific assessment outcomes and employ various concept mapping tasks. The system provides different informative, tutoring and reflective feedback components, tailored to learners' individual characteristics and needs by using weight and error categories. The level of performance is represented by Gogoulou's taxonomy (Gogoulou *et al.*, 2004). COMPASS supports the identification and the qualitative analysis of errors presented on the learner's map. The results are further utilized for the qualitative diagnosis of the learner's knowledge and the quantitative estimation of the learner's knowledge level on the activity, according to assessment criteria defined by the teacher. Aiming to individualise feedback, COMPASS provides to the learner different informative and tutoring feedback components, tailored to the learner's knowledge level, preferences and interaction behavior. The feedback provided aims to stimulate the learners to reconsider their beliefs, reflect on them and reconstruct/refine their knowledge structure. In summary, this system still has the same problems as INSPIRE.

4. SIETTE (Spanish translation of Intelligent Evaluation System using Tests for TeleEducation)

SIETTE is a web-based tool to assist teachers and instructors in the assessment process. The system can be used in two different ways. First, teachers can use it to develop the tests that are defined by their topics, questions, parameters, and specifications. Second, learners can use it to take the tests that are automatically generated according to the specifications, and adapted to the learner's knowledge level. As the learner answers the items, the new ability level is computed and the next question selected, until the stopping criterion is met. Question selection is based on a function that estimates the probability of a correct answer by using Item Response Theory, leading to an estimation of the learner's knowledge level. This system has the problem with estimating learner's knowledge level of each topic in each test.

5. Athena QTI

Athena QTI is a web-based adaptive assessment authoring system. It focuses on presenting the functionality of the authoring environment and the tool's conformance to the IMS QTI specification. Assessment items are selected and presented to the learner according to a set of rules that the author creates by using a form of IF-THEN rules. The condition refers to user model information, and the action referring to the result of changing in the assessment. The authors claim that this system is given the flexibility of expressing a teacher's didactical philosophy and methods through the creation of an appropriate rule. That is better than SIETTE, which uses functions to estimate any parameter, such as estimation of the probability of a correct answer and learner's knowledge level. The highest probability question will be posed. This system is still in its infancy. There are some problems with estimating and representing learner's knowledge level, formal testing with the real users, and evaluation of the adaptation features of the assessments created.

6. CosyQTI

The CosyQTI tool supports the authoring process and presentation of personalized and adaptive web-based assessment. Adaptation in work means two things: the generation of a dynamic sequence of questions depending on a learner's responses and estimation of learner knowledge level, and the adaptation of the content of web-based tests to the learner's access device characteristics such as screen dimensions, storage capacity and processing power. The adaptation will be provided by using a form of the IF-THEN rule's trigger point which is a point for activation. In addition, CosyQTI allows learners to access parts of their model to raise learners' awareness of their developing knowledge, and difficulties of the learning process, leading to enhanced learning. IMS QTI, IMS LIP and IEEE PAPI specifications are implemented in this system to provide interoperability. This system has not been tested in full in real classroom environments. There are still some problems with estimating and representing learner's knowledge level and formal testing within real environments.

The comparisons between these systems are provided in Table 3-1 and Table 3-2.

System	Personal-reader	INSPIRE	COMPASS	SIETTE
Standard and technology	IEEE LOM, IMS QIT, IEEE PAPI, and Semantic web technologies	HTML and ASP	HTML and ASP	HTML and JAVA
Defined rule	Using TRIPLE language and First Order Logic	Using Item Response Theory	Using weight and the error categories	Using Item Response Theory
Using level of performance	No	Yes	Yes	No
Adaptive assessment process	Yes	No	No	Yes
Adapting learning presentation	No	Yes	Yes	No
Adaptive content : Presentation	Yes	Yes	No	No
Adaptive content : Link	Yes	Yes	No	No
Adaptive feedback	Yes	Yes	Yes	No
Diagnostic assessment	Yes	Yes	Yes	No
Formative assessment	Yes	Yes	No	No
Summative assessment	Yes	No	Yes	Yes
Select goal	No	Yes	Yes	No
Select lesson	Yes	No	No	No
Reusability	Yes	No	No	No
learning style	No	Yes	Yes	No
Knowledge representation	Learning Object	Learning Concept	Concept Map	A curriculum-based structure and topics and questions

Table 3-1 Comparison of adaptive assessment system

System	AthenaQTI	CosyQTI
Standard and technology	IMS QTI	IMS QTI, IMS LIP and IEEE PAPI
Defined rule	Using the form of IF-THEN rules	Using the form of IF-THEN rules
Adaptive assessment process	Yes	Yes
Adapting learning presentation	No	Yes
Adaptive content : Presentation	No	No
Adaptive content : Link	No	No
Adaptive feedback	Yes	No
Adaptive testing	No	No
Adaptive Question	Yes	Yes
Diagnostic assessment	No	No
Formative assessment	No	No
Summative assessment	Yes	Yes
Select goal	No	No
Select lesson	No	No
Reusability	Yes	Yes
learning style	Yes	Yes
Knowledge representation	Not mentioned	Not mentioned

Table 3-2 Comparison of adaptive assessment system (continued)

3.7 Some problems of adaptive assessment

According to Table 3-1 and Table 3-2, some problems of adaptive assessment are found as follows (Sitthisak, Gilbert and Davis, 2007).

1. **Inconsistency arising from adaptive assessment systems estimating the learner's knowledge level differently**

There are many systems using the number of questions answered correctly, and the difficulty level of answered questions, in order to estimate the ability or knowledge level of each learner, such as Personal-reader (Cheniti-Belcadhi and Braham, 2004), INSPIRE (Papanikolaou *et al.*, 2002), and COMPASS (Gouli *et al.*, 2004). Each system classifies ability or knowledge level and difficulty level of assessment using different approaches and techniques. Most adaptive assessment systems do not easily permit reuse or allow the exchange of a learner's knowledge level between learning management systems (Cheniti-Belcadhi and Braham, 2004; De Bra, Aroyo and

Chepegin, 2004). This causes interoperability and reusability problems if the learner's knowledge level in one system needs to be used in other systems.

2. **The limited affordance offered, in what is a multidimensional problem, by simply using a numerical value to match a learner's knowledge level**

There are many well-known theories for selecting questions in order to match a learner's knowledge level, such as granularity hierarchies, Bayesian nets, and Item Response Theory (IRT) (Collins, Greer and Huang, 1996). These theories have assumptions concerning the mathematical relationship between abilities and item responses. A numerical value from these theories may be appropriate to decide who the best learner is, but an evaluation of education intends to assess the learners' readiness for further learning (Falmagne *et al.*, 2003). Therefore, selecting a question in adaptive assessment should be multidimensional.

3. **The dependency of existing adaptive assessment systems on specific knowledge domains in supporting lifelong learning**

In most cases, adaptive assessment systems are developed for a specific knowledge domain using particular rules and assessments without possibility for knowledge reuse, for example AthenaQTI (Tzanavari, Retalis and Pastellis, 2004) and CosyQTI (Lalos, Retalis and Psaromiligkos, 2005). There are many difficulties for updating rules, content and assessment of those systems. Most adaptive assessment systems lack reusability as there is no standard to combine their different knowledge domains with assessments and learned capabilities (Cheniti-Belcadhi and Braham, 2004). This highlights the problem of supporting lifelong learning assessment. Lifelong learning is about "acquiring and updating all kinds of abilities, interests, knowledge and qualifications in order to promote the development of knowledge and competences throughout life" (The European Commission, 2006).

3.8 Summary

This chapter first introduced the concept of effective questions and a discussion about the application of automatic generating question systems. This was followed by a review of the technologies of adaptive hypertext and adaptive assessment. Presentation adaptation and navigation adaptation were discussed. Subsequent sections have explored the capability of adaptive assessment applications, presented the comparison of adaptive assessment systems, and highlighted some of the problems faced by the authors of adaptive assessment application. These problems are: inconsistency arising from adaptive assessment systems estimating the learner's knowledge level differently, the limited affordance offered; in what is a multidimensional problem, by simply using a numerical value to match a learner's knowledge level, and the dependency of existing adaptive assessment systems on specific knowledge domains in supporting lifelong learning. The problems may emerge, not only from adaptive hypertext or adaptive assessment themselves, but also from an awareness of using pedagogy and the potential of learning design.

The following Chapter presents knowledge definition, knowledge engineering process, and a description of the Semantic Web in order to understand how to collect and represent learners' knowledge.

Chapter 4 Knowledge Engineering

4.1 Introduction

This chapter considers the question "What is knowledge engineering?". A detailed description of the Semantic Web and ontologies is discussed. Finally, issues of applying ontologies in education and e-learning are presented.

4.2 The definition of knowledge

The term "knowledge" is defined in a complex of several related ideas as follows. In Smith (1996), "Knowledge consists of symbols, the relationships between them and rules or procedures for manipulating them." In Lukose (1996), "Knowledge is more than a static encoding of facts, it also includes the ability to use those facts in interacting with the world."

To conclude, it seems that knowledge is a kind of collection of related information which can be connected when collected; it becomes knowledge (Gadomski, 2001). The combined nature of many such connections provides greater meaning for greater use and value, as illustrated in the development of the Advance Knowledge Technologies (AKT) project. The AKT project involves five universities from the UK, and aims to develop and extend a range of technologies providing integrated methods and services for the capture, modelling, retrieval, publishing, reuse, and maintenance of knowledge (AKT Manifesto, 2000).

In the following section, the issues of how to collect and represent knowledge are focused on the knowledge engineering process.

4.3 Knowledge engineering process

According to the conceptual nature of knowledge, knowledge is dynamic and changes with time. The process of collecting and representing knowledge is a crucial task. The person who has the knowledge may not be able to express that knowledge in an English-like form. A key process in the development phase of knowledge engineering is the acquisition and representation of knowledge.

4.3.1 Knowledge acquisition

Knowledge acquisition is the process of eliciting, analysing, and modelling the pattern of knowledge underlying some subject matter for knowledge engineering (Sowa, 2000). There are many methods used for acquiring, analysing and modelling knowledge, such as gaining the knowledge from a printed source, or an interview, and a discussion or questionnaires (Smith, 1996). In this study, subject matter content, capability taxonomy, and competence, were collected from core textbooks and websites of course syllabus in INFO 1013 IT modelling course and INFO 2007 Systems Analysis and Design course at the University of Southampton.

4.3.2 Knowledge representation

There are many experts in this area who have summarised the basic ideas about knowledge representation as follows.

Davis, Schrobe, and Szolovits (1993) wrote about five basic principles in knowledge representation. First, a knowledge representation is a surrogate. Physical objects, events, and relationship cannot be stored directly in a computer so they are represented and stored in some suitable form to allow for subsequent computer processing. Second, a knowledge representation is a set of ontological commitments. Ontology determines the categories of knowledge. Third, a knowledge representation is a fragmentary theory of intelligent reasoning. The behaviour and interactions of the things in a domain must be represented to support reasoning. Next, a knowledge representation is a medium for efficient computation. A representation provides for processing information on the available computing equipment. Finally, a knowledge representation is a medium of human expression. A good knowledge representation language should facilitate to express knowledge in term of everyday words, phrases, diagrams, etc.

Sowa (2000) concluded that "Knowledge representation is a multidisciplinary subject applied from logic, ontology and computation field. Logic provides the formal structure and rules of inference. Ontology defines the kinds of things that exist in the application domain. Computation supports the applications for implementing the logic and ontology in computer programs."

One aspect of knowledge representation schemes is the semantic network or semantic net. Semantic networks allow the meaning or semantics of objects to be represented as hierarchical relationships between objects (Gonzalez and Dankel, 1993; Smith, 1996).

A semantic network is made up of a number of nodes which represent objects and detailed information about those objects. Each node is connected by links which represent the relationships between the objects. In the semantic network, adding nodes and links to a semantic network does not change the structure of any existing sub-graph (Sowa, 2000). Inheritance is a property of the semantic networks which refers to the ability of one node to inherit characteristics from other nodes in the network (Gonzalez and Dankel, 1993).

In this study, a domain expert expressed domain content, the capability taxonomy, and competence in an English-like form. A knowledge engineer represented these elements in the form of a semantic network, and then transformed them into an ontology.

4.4 Semantic web

In this century, it is clear web technology has developed rapidly to support the essential property of the World Wide Web which is universality. At present, representing the meaning of objects on the web to allow machine-accessibility is the main obstacle to supporting web users (Antoniou and Harmelen, 2004). A semantic network has been applied to the web in order to allow any existing knowledge representation system to be exported onto the web. This is called the Semantic Web.

Berners-Lee, inventor of the Web (Berners-Lee, Hendler and Lassila, 2001), defined "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation".

The power of the Semantic Web is that machines become much better able to process and understand the data that they merely display at present, such as data and rules for reasoning. The basic technologies of the Semantic Web are as follows (Antoniou and Harmelen, 2004).

1. Explicit metadata

   The Semantic Web includes the metadata which capture part of the meaning of data. The metadata is more easily processable by machines.

2. Ontologies

   An ontology is a core component in the Semantic Web, and is an explicit and formal specification describing the main concepts of a domain and providing a shared understanding of a domain. More detail of ontologies are discussed in the following section.

3. Agents

   Agents are computer programs that work autonomously and proactively for the person. However, they receive some tasks from a person, make certain choices, and give answers to the user.

The Semantic Web has the potential to increase the effectiveness of educational functions according to three fundamental affordances which are (1) the semantic conceptualisation and ontologies, (2) common standardised communication syntax, and (3) large-scale service-based integration of educational content and functionality provision and usage (Aroyo and Dicheva, 2004). Anderson and Whitelock (2004) also support this view that the vision of the educational semantic web is based on the capacity for effective information storage and retrieval, the capacity for non-human autonomous agents to augment the learning and information retrieval, and the capacity of the internet to support, extend, and expand communications capabilities of humans.

4.4.1 Semantic Web architecture

To express meaning of a resource on the web, the layers of the Semantic Web architecture are presented in Figure 4-1 (Berners-Lee, 2000). The layers of the Semantic Web are described as follows.

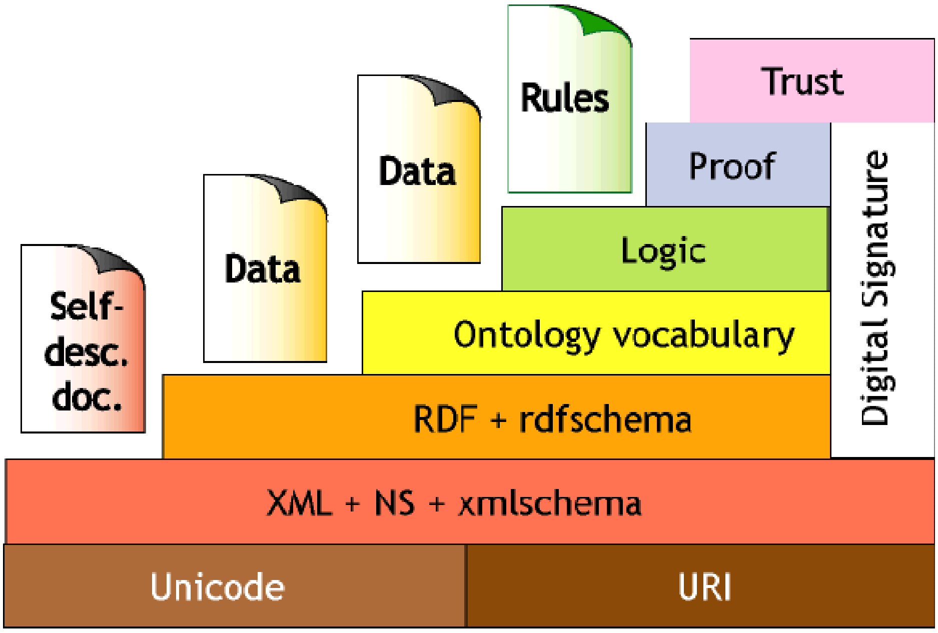


Figure 4-1 Semantic Web architecture (Berners-Lee, 2000)

The Unicode and URI layer represents a global naming scheme. A URI is simply a web identifier. Every data object and every data scheme/model in the Semantic Web must have a unique URI.

The XML layer represents data to accomplish the exchange of data between web applications. XML (eXtensible Markup Language) allows everyone to create their own tags or arbitrary structure to their document but the meaning of the structure is not mentioned. So, XML is not the solution for adding semantics to the web resources.

The RDF layer represents the meaning of data. The Resource Description Framework (RDF) is a specification that defines a model for adding semantics to a document. RDF represents the description of document in form of object-attribute-value triples, that is: an object *O* has an attribute *A* with the value *V*.

RDF Schema (RDFS) is a semantic extension of RDF. Basic groups of related documents and the relationships between those documents are expressible in RDF Schema. However, RDFS is limited since it models a document in a hierarchical structure which is not capable of representing disjointness of groups of related documents.

The ontology layer uses the Web Ontology Language (OWL) to represent the formal common agreement about meaning of data. OWL adds more vocabulary to describe properties and classes than RDF or RDF Schema. In addition, it can describe relations between classes such as disjointness, cardinality, and characteristics of properties. OWL is designed for use by applications that need to process the information contained in documents. OWL has three sublanguages: OWL Lite, OWL DL, and OWL Full. OWL Lite supports a classification hierarchy and simple constraint features such as thesauri and taxonomies. OWL Description Logic (DL) supports the maximum expressiveness with computational completeness. OWL Full is fully compatible with any reasoning software. It supports maximum expressiveness with no computational guarantees.

In this study, the ontology was based on OWL-Lite which was sufficiently expressive to describe the subject matter hierarchy and provides for higher performance reasoning.

4.4.2 Ontologies

Different definitions of ontologies are based on different backgrounds and interests. Berners-Lee, Hendler and Lassila (2001) defined an area of artificial-intelligence and web thus: "Ontology is a document or file that formally defines the relations among terms. The most typical kind of ontology for the web has a taxonomy and a set of inference rules." In the context of knowledge representation, Sowa (2000), defined "An ontology is a catalogue of the types of things that are assumed to exist in a domain of interest *D* from the perspective of a person who uses a language *L* for the purpose of talking about *D*." In the context of philosophy, Gruber (1993) defined "an ontology" as "an explicit specification of conceptualisation". An ontology was defined by a set of representational terms which associated the name of entities in the universe of discourse with human-readable objects.

According to the W3C (2004a), "An ontology defines the terms used to describe and represent an area of knowledge. Ontologies are used by people, databases, and applications that need to share domain information. Ontologies include computer-usable definitions of basic concepts in the domain and the relationships among them … They encode knowledge in a domain and also knowledge that spans domains. In this way, they make that knowledge reusable." This definition was defined on the context of the Semantic Web. It means that an ontology is representation of knowledge and information about a particular domain that is both machine- and human-readable and portable across domains.

Based on the reviewing of what an ontology stands for, Kalfoglou (2001) summarises "the explicit representation of a shared understanding of the important concepts in some domain of interest…" He also proposed the criteria of design ontologies which are clarity, coherence, extendibility, minimal encoding bias, and minimal ontological commitment.

- Clarity refers to the minimal ontological ambiguity. All definitions should be communicated effectively.
- Coherence means that the ontology should be internally and logically consistent.
- Extendibility means that adding new terms in the existing ontology should be flexible without the revision of existing definitions.
- Encoding bias should be minimised when the representation is made purely for the convenience of notation or implementation.
- Minimal ontological commitment means that the ontology should allow the parties committed to the ontology freedom to specialise and instantiate the ontology as required.

The ontologies in this study adhered to these criteria of ontology design to ensure that the ontologies could be reused.

4.4.3 Building ontologies

The construction of an ontology is a time-consuming and complex task. Two methodological approaches have been proposed and used as follows (Dicheva *et al.*, 2005; Gómez-Pérez and Manzano-Macho, 2004). Manual ontology generation is time-consuming and focused on research problems or related to the ontology engineering process. Automatic and semi-automatic ontology generation and extraction approach, uses different kinds of methods such as linguistic techniques, statistical techniques, and machine learning algorithms. It is more appropriate to speed the creation of new ontologies.

Although the method of semi-automatic generation is more appropriate for speeding up the process of ontology generation, this study used the manual technique to build its domain ontologies.

4.4.4 Ontologies in education and e-learning

Two types of ontology dealing with many ontology-based applications are domain ontology and structure ontology (Dicheva *et al.*, 2005). Domain ontology represents the basic concepts of the domain under consideration along with their interrelations and basic properties. Structure ontology defines the logical structure of the content. It is generally subjective and depends greatly on the goals of the ontology application.

In this study, the implemented ontologies are domain, not structure, ontologies which are compliant with a hierarchy of knowledge (Merrill, 1983). These ontologies used a controlled vocabulary from Simple Knowledge Organisation System (SKOS) (W3C, 2005).

Ontologies have been considered as a knowledge base component (Dicheva *et al.*, 2005), supporting the presentation and delivery of course material and for assisting and assessing students (JISC, 2004). This section focuses on some applications in the learning technologies prospect according to the areas of research.

In the area of learning objects, the approach of using ontologies and the Semantic Web to represent content and structure of the learning materials is proposed (Jovanovic, Gasevic and Devedzic, 2006). Applications are implemented as a learning web application to generate content, which is semantically personalised to the learner's goals, preferences, and learning styles (Stojanovic, Staab and Studer, 2001). This application intends to provide a more comfortable search and navigation through the learning material.

In the area of educational web portals, Dicheva, Sosnovsky, Gavrilova, and Brusilovsky (2005) developed the Ontologies for Education (O4E) Web Portal for publishing the created ontology and serving as a point of access to the relevant online information. Ontoportal was an ontological hypertext framework for building educational web portals based on a simple domain ontology (Woukeu *et al.*, 2003).

In the area of Web-Based Educational System (WBES), Semantic Web technologies were employed in WBES to enhance adaptation and flexibility (Aroyo and Dicheva, 2004). Topic Maps for Learning (TM4L) (Dicheva, Dichev and Wang, 2005) and AIMS (Aroyo and Dicheva, 2001) involved exploring the domain ontology and searching the repository for information related to a specific task.

4.5 Summary

This chapter has given the field of knowledge engineering. The processes of knowledge engineering that a knowledge engineer will go through have been discussed. The importance of the Semantic Web, and why it is appropriate for education and e-learning, has been covered. Finally, the ontological approach and application in education and e-learning have been presented. The following chapter presents competency definitions in order to support representation of competency, based on ontologies. In addition, the existing competency standards and e-learning specifications are introduced to investigate the existing problems of these specifications.

Chapter 5 Competency models and e-learning

5.1 Introduction

Today's societies place challenging demands on individuals who need to acquire key competencies. Defining such competencies can improve assessment of how well prepared learners are, as well as identify overarching goals for education systems and lifelong learning. This chapter first reviews competency definitions and existing competency standards: IMS Reusable Definition of Competency or Educational Objective (RDCEO) specification and HR-XML competency standard. Finally, this chapter provides an introduction to e-learning standards and specifications: IMS Content Packaging (CP), IMS Question and Test Interoperability (QTI), IMS Learning Design (LD), and SCORM. These specifications of competency and e-learning are introduced to explore the potential of competency as an emergent pedagogy.

5.2 Competency definition

In this section, competency and competence definitions are presented as follows. The HR-XML Consortium defined competency as "A specific, identifiable, definable, and measurable knowledge, skill, ability and/or other deployment-related characteristic (e.g. attitude, behaviour, physical ability) which a human resource may possess and which is necessary for, or material to, the performance of an activity within a specific business context." (HR-XML Consortium, 2006).

Competence may be defined as the integrated application of knowledge, skills, values, experience, contacts, external knowledge resources and tools to solve a problem, to perform an activity, or to handle a situation (Friensen and Anderson, 2004; Sandberg, 2000).

In a paper which described the professions, competence and informal learning, Cheetam and Chivers (2006) defined the concept of competence as relating to three different dimensions: a personal competence, a context, and the proficiency level of a person with respect to a context.

The competency and competence definitions given above may not make significant distinctions relevant to this thesis. Both concepts of "competency" and "competence" are adopted in this study, which will be called "competency" in the rest of this thesis (two or more will be called "competencies").

A competency model should support storing, organising and sharing of achieved, current, and intended performance data relating to all aspects of education and training in a persistent and standard way, so as to ensure that learners can find learning activities that fit and improve their acquired competencies.

5.3 Competency standards

According to competency definitions, how these definitions conform to the existing meta-data model are investigated in the IMS RDCEO specification (IMS RDCEO, 2002) and the HR-XML Consortium competencies schema (HR-XML Consortium, 2006).

5.3.1 IMS RDCEO

IMS Global Learning Project has developed a Reusable Competency Definitions Information Model (RDCEO), which may be useful in capturing definitional and descriptive information about competencies.

The IMS RDCEO specification presents competency information in five categories: *Identifier* holds a globally unique label to reference the competency in any other system. *Title* holds a short human-readable name for the competency. *Description* holds a human readable description of the competency. *Definition* holds a structure for including an arbitrary collection of statements that determine a competency. It provides for model sources and statements to express different types of competencies and competency models. *Metadata* holds a meta-data record conforming to IEEE LOM including additional information such as author and creation date (IMS RDCEO, 2002).

Because of the unstructured textual definitions in RDCEO, the specification allows only the description of the proficiency level via the title element. This is a problem for having the level of the described competency separate from its narrative description. In the paper which mentioned adopting e-learning standards in health care, the problem of linking competency to the content of learning materials because of the RDCEO unstructured textual descriptions are discussed (Hersh *et al.*, 2006). Finally, there remain problems with the grading scale of a competency, the success threshold of a competency, and the structure of complex competencies within RDCEO (Karampiperis, Sampson and Fytros, 2006).

5.3.2 HR-XML

In the area of emerging standards, the HR-XML consortium was established to create an XML schema for providing standardized and practical means to exchange information about competencies within a variety of business contexts. The aim of this outline is to develop the competencies schema in order to allow the capture of information about evidence used to substantiate a competency, and ratings and weights that can be used to rank, compare, and otherwise evaluate the sufficiency or desirability of competency (HR-XML Consortium, 2006).

The HR-XML consortium creates this competency standard schema based on its competency's definition. The HR-XML Consortium define competency as "A specific, identifiable, definable, and measurable knowledge, skill, ability and/or other deployment-related characteristic (e.g. attitude, behaviour, physical ability) which a human resource may possess and which is necessary for, or material to, the performance of an activity within a specific business context."

HR-XML presents competency information in nine categories: *Name* holds the competency name. *Description* holds additional information about the competency. *Required* holds a boolean to indicate whether competency evidence is mandatory. *CompetencyId* holds an identification code to identify or classify the competency. *TaxonomyId* holds a code to reference taxonomies, for example Bloom's cognitive domain. *CompetencyEvidence* holds information about the existence, sufficiency, or level of the competency, such as test results or certificates. *CompetencyWeight* provides information on the relative importance of the competency. *Competency* allows for decomposition into component competencies, such that the competency may include other competencies. *UserArea* holds job-related information, such as the job position.

The HR-XML competency standard is focused on helping an organization improve communication across its HR activities enhancing recruiting systems, rather than on improving the use of competency information in education or training.

5.4 E-learning standards

E-learning standards are adapted from multiple sources to provide a comprehensive suite of e-learning capabilities that enable interoperability, accessibility and reusability of web-based learning content (IMS LD, 2003). There are several existing specifications and standards such as IMS Content Packaging (CP), IMS Question and Test Interoperability (QTI), IMS Learning Design (LD), and SCORM. Their details are presented below.

5.4.1 IMS Content Packaging

In the context of e-learning, content refers to items such as blocks of text, pictures, diagrams, animations, tests and answers, learner information, resource information, collaborative tools, etc. (Burgos and Griffiths, 2005). These contents, which are part of the e-learning material in the teaching-learning process, need to be packaged. The reasons for packaging are to form a coherent e-learning course, to be stored in digital repositories in order to be made accessible to many learners, dispersed areas and to address the reusability of the course materials (Chew, 2001). Hence, IMS CP Specifications were designed containing necessary meta-data information to find the relevant content for the learner, to move contents from one location to another, and to interoperate in different learning management systems by use of XML (IMS Content Packaging, 2005).

5.4.2 IMS Question and Test Interoperability

The IMS QTI specification, defined by the IMS Global Learning Consortium, is part of the same family of specifications as IMS Learning Design. IMS QTI specification is established to describe a data model for representing question and test data, as well as their corresponding result reports. In addition, this specification has been designed to support both interoperability and innovation (IMS QTI, 2006). It describes the basic structure that is necessary to represent questions (AssessmentItem) and test of evaluations (AssessmentTest). Moreover, this specification enables the exchange of these items; test and results data between authoring tools, item banks, test constructional tools, as well as learning systems and assessment delivery systems. QTI version 2.0 processing is illustrated in Figure 5-1.

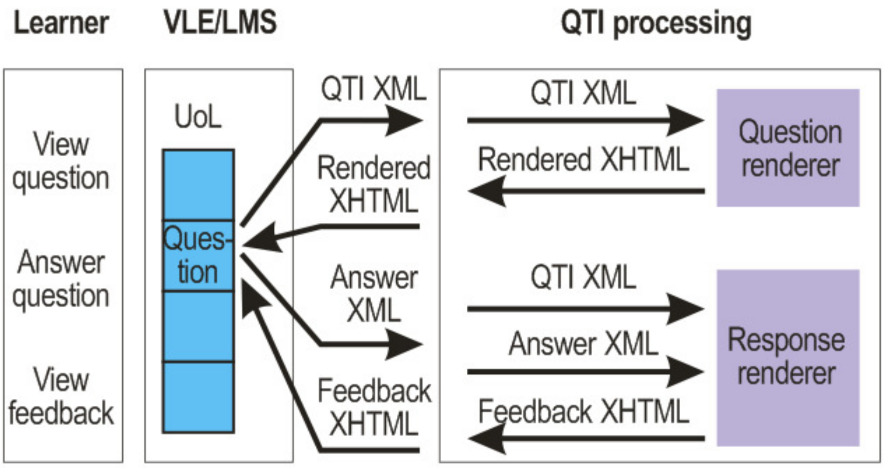


Figure 5-1 QTI version 2.0 processing (Wills *et al.*, 2006)

When a learner accesses a Virtual Learning Environment or Learning Management System (VLE/LMS) to view and respond to a QTI question, the system initially sends a QTI XML file to a QTI processing service where a Question renderer renders the question, the rendered question is sent back to the VLE/LMS for display to the learner. The learner's answer is sent to a QTI Response renderer which marks the answer and provides feedback. The rendered feedback is sent back to the VLE/LMS for display to the learner.

5.4.3 IMS Learning Design

IMS LD starts from the position that learning is different from content consumption and that learning comes from being active. It recognises that learning happens when learners cooperate to solve problems in social and work situations (Griffiths and Blat, 2005). IMS LD is based on the following principles: in a learning process each person has a role (learner or teacher) and seeks to obtain results by carrying out learning activities and/or support within an environment. The major concept of the IMS LD is the method which allows the coordination of activities of each role in the associated environment to achieve learning objectives according to prerequisites. The learning process is modelled on a theatrical play from a structural point of view (IMS LD, 2003).

5.4.4 Integrating IMS LD and IMS QTI

There are many tools for implementing e-learning standards including RELOAD (Reuseable eLearning Object Authoring and Delivery) editor, RELOAD learning design editor and CopperCore, discussed below.

The RELOAD editor (Reload, 2004) is a content package and metadata editor. It can package electronic content such as web pages, images, etc., and describe them ready for storage in content repositories such as JORUM. At present, the RELOAD editor supports V1.1.3 of the IMS Content Packing specification and V1.2.2 of IMS Metadata specification, as well as V1.2 of the SCORM package.

The RELOAD learning design editor (Reload, 2006) based on the IMS Learning Design specification, supports the creation of re-useable "Pedagogical Templates", allowing the user to define a set of learning objectives, activities, and learning environments.

The CopperCore (CopperCore, 2005) is an IMS Learning Design engine, which supports all three levels of IMS Learning Design. It provides the core functionality of interactions between the various roles, resources and activities. However, other developers can implement an interface on top of it.

These tools were used to implement a Thai tea unit of learning (Sitthisak *et al.*, 2007b), in order to explore how to implement learning design in a general way. This covers the teaching-learning process such as interactions between teachers and students, collaborative learning, adaptive learning and personalisation, teacher monitoring and conformance to IMS QTI and SCORM.

The IMS QTI specification can be considered as an integrative layer in implementing IMS LD Unit of Learning. However, this implementation showed some shortcomings, such as ineffective interactivity, difficulty of learning design coding, inflexibility, and poor reusability for the group study implementation. Instructional designers should consider this issue when integrating IMS QTI items within an IMS LD Unit of Learning.

5.4.5 SCORM

SCORM (Sharable Content Reference Model) is a suite of technical standards that enable web-based learning systems to find, import, share, reuse, and export learning content in a standardized way. In addition, SCORM targets the web as a primary medium for delivering instructions. It does so under the assumption that anything delivered by the web can be easily used in other instructional settings that make fewer demands on accessibility and network communications. It describes two common ways: how to create web-based learning content that can be delivered and tracked by learning management systems, and what a learning management system must do to properly deliver and track SCORM compliant learning content (SCORM, 2004).

5.5 Summary

A competency is more than just knowledge and skills. It involves the ability to meet complex demands in a particular context. There are two important competency standards: the IMS RDCEO specification and the HR-XML Consortium competencies schema. However, these standards suffer some shortcomings and still miss an important point of competency relations and tools. Finally, Four e-learning standards and specifications: IMS CP, IMS LD, IMS QTI, and SCORM have been presented. The following chapter outlines the development of the competency model to improve the use of competency information in education.

Chapter 6 An improved competency model

6.1 Introduction

The IMS RDCEO specification still has problems with: the level of the competency described separated from its narrative description; the grading scale of a competency; the success threshold of a competency; and the structure of complex competencies within RDCEO. One of the problems with the HR-XML competency standard is that, in focusing on helping an organization improve communication across its HR activities by enhancing recruiting systems, it does not address improving the use of competency information in education and training. Solving these problems results in the following proposed competency model, reflecting all relevant features of the learner's behaviour and their knowledge, skills, and attitudes that affect their learning and performance. The following sections analyse the criteria for a good competency model, and requirements for competency standards.

6.2 The criteria for competency models

Competency is defined as the integrated application of knowledge, skills, values, experience, contacts, external knowledge, resources and tools to solve a problem, to perform an activity, or to handle a situation (Friensen and Anderson, 2004; Sandberg, 2000). The criteria for competency models are as follows (Sitthisak, Gilbert and Davis, 2007).

First, a competency model should maintain a rich data structure for description, comprehensive reference, and exchange, to support a learner's competency profile throughout their life. In order to assess learned capability and perform competency gap analysis, the model should support recording competency achievements and the attainment of intended learning outcomes.

Second, meeting personal needs requires highly flexible competency-based learning. Many learners have different roles, proficiencies, preferences, abilities and backgrounds. A good competency model should support such personalisation.

Third, monitoring and recording a learner's competency is important for selecting suitable questions in an adaptive assessment system. Mechanisms for selecting questions are based on learning progress and decisions about the further direction of the learning process. A good competency model should support straightforward transformations between competency statements and assessment of such competencies.

Fourth, competency should be concerned with specific, identifiable and measurable behaviours (Draganidis and Mentzas, 2006). It enables the creation of assessments by transforming learned capabilities to question statements. This supports the automatic collection and expression of assessment for individual and group competencies.

6.3 Comparison of competency standards

In this section, the possible requirements are listed for describing competencies based on an analysis of the general structure of existing competency standards and competency ontologies (Draganidis and Mentzas, 2006; Schmidt and Kunzmann, 2006; Trichet and Leclère, 2003). The requirements are classified into nine categories, where each is divided into sub categories. The requirements list is general and captures the types of information modelled in existing standards, rather than defining a canonical set of properties.

1. Description: the general description of the competency.
2. Type: type of trait that represents an aspect of the competency such as knowledge, skill, attitude, and so on.
3. Relationship: relationship to other competencies such as "part-of", "child competency", and "parent competency".
4. Proficiency level: a measurement of the degree to which the competency has been achieved.
5. Measurement scale: a scale that relates to proficiency level and weight.

6. Taxonomy: a taxonomy reference for structuring competency data.
7. Evidence: facts or indicators about the achievement of a competency, such as test results and certificates.
8. Tools: any tool(s) required to support reaching the competency.
9. User area: Other data, such the description of a job position.

A comparison of two competency standards according to these requirements, is shown in Table 6-1 (Sitthisak *et al.*, 2007a).

Categories	Sub-categories	IMS RDCEO	HR-XML
Competency description		■	■
Competency type	Knowledge	◘	◘
	Skill	◘	◘
	Attitudes	◘	◘
Competency relationship		◘	□
Proficiency level		◘	■
Measurement scale		◘	■
Taxonomy		◘	■
Evidence		◘	■
Tools		◘	□
User area		◘	■

Support: '■' = full, '◘' = partial, '□' = none

Table 6-1 A comparison of the capabilities of competency standards

First, IMS RDCEO provides a flexible definition of competency using unstructured textual definitions. Often a less precise definition is very useful, especially when dealing with competency data from different communities of practice. However, this leads to shortcomings in domain definition, ontology use, the ability to compare competency data between different communities, and the tracking of the knowledge state of the learner.

Second, HR-XML addresses some shortcomings of RDCEO, as illustrated in Table 6-1. However, it still misses the important points of competency relations and tools. Although HR-XML provides for competencies to be composed of other competencies, it does not have an element referring to the competency relation. This may cause selection problems. For example, in a competency hierarchy, it should be possible to specify which elements of the competency hierarchy are mandatory and which are optional.

These existing e-learning competency standards, however, are not able to accommodate complicated competencies, link competencies adequately, support comparisons of competency data between different communities, or support tracking of the knowledge state of the learner.

An improved competency model, named COMpetence-Based learner knowledge for personalized Assessment (COMBA), is proposed in this study (Sitthisak, Gilbert and Davis, 2008c). COMBA is informed by the results of comparing the competency standards against the desired taxonomy of competence and this point will be discussed below.

6.4 An improved competency model

The issue of how to represent competency as a rich data structure is focused on supporting collaboration between different communities and the tracking of the knowledge state of the learner. The same competencies may appear in more than one place in the competency hierarchy. Thus, it makes sense to capture the data model of those competencies in some reusable form, so they have to be defined only once. The improved competency model is represented in Figure 6-1. The heart of this model is the treatment of knowledge, not as possession, but as a contextualized multidimensional space of either actual or potential capability.

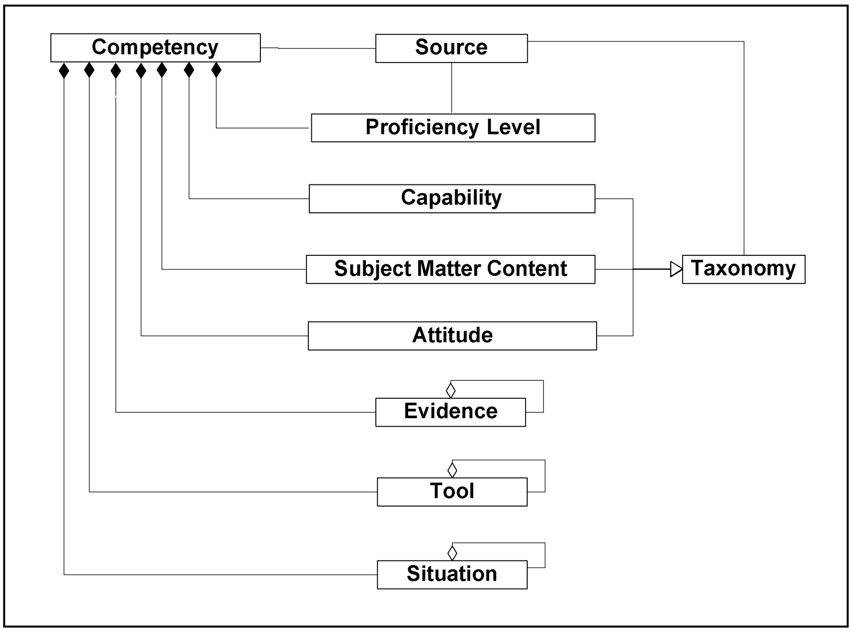


Figure 6-1 Competency model

A competency involves a capability associated with subject matter content, a proficiency level, evidence, any required tools, and definition of the situation which contextualises the competency. Each competency, proficiency level, capability, attitude, and subject matter content, has a source which refers to these elements.

Capability is behaviour that can be observed, based on a domain taxonomy of learning such as Bloom's (Bloom and Krathwohl, 1956), Gagné's Nine Areas of Skill (Gagne, 1970), or Merrill's Cognitive Domain (Merrill, 1999).

Subject matter content is the subject domain of what the learner can do by the end of course.

Attitude is the way in which a learner exhibits their knowledge and skill, perhaps categorised using a version of Krathwohl's taxonomy (Krathwohl and Anderson, 2002).

Proficiency level indicates the level of proficiency that learners should or do possess of a particular competency.

The competency *evidence* substantiates the existence, sufficiency, or level of the competency, and might include test results, reports, evaluation, certificates, or licenses. External knowledge resources and *tools* support and promote the problem solving, activity performance or situation handling of the competency. The *situation* identifies the particular context and conditions of the competency, for example, its time limit.

The proposed competency model involves three important elements:

- an orientation towards, and focus upon, activity-based teaching and learning
- the identification and integration of appropriate subject matter content within a broader teaching and learning context, represented by a hierarchy of competencies
- the identification of the assessment that would demonstrate successful teaching and learning has been accomplished.

Competency modelling needs multiple hierarchies for cross-reference between disciplines (Kunzmann, 2006). In the proposed competency model, linkages between competencies within a competency tree are separated from the competency records themselves.

6.4.1 Instructional theory

In this study, the competency is aligned with the instructional content and ability matrix of intended learning outcomes, as illustrated in Table 2-1 (Chapter 2). The combination of Bloom's taxonomy and Merrill's analysis of subject content are presented in the competency model. This model gives a competency tree structure and meaning to an instructional design problem, enabling the instructional designers to discuss their design task with an appearance of understanding. Sequences of questions are aligned with the learner's level of knowledge based on the concept of a hierarchy of knowledge and their cognitive ability, in order to use questioning more effectively as a pedagogical strategy (Gilbert and Gale, 2007). As mentioned in Chapter 3, the criteria for effective questions are also considered in the generation of questions.

6.4.2 Assessment systems

Currently, many systems such as Problets (Dancik and Kumar, 2003), ILE (Cristea and Tuduce, 2005), QuizPACK (Brusilovsky and Sosnovsky, 2005), and Jeliot 3 (Myller, 2007) offer remarkable automatic generation of questions and adaptation of questions, but only for specific domains, and they lack integration, interoperability, portability, and reusability with other systems and environments. In addition, such systems are difficult to use in e-Learning systems, particularly in assessment systems. For example, consistency checking, assessing differences in knowledge levels, and comparing achievement in related domains remain essentially impractical. In addition, the main difficulty for learners undertaking e-Assessment may be that the number of available questions is insufficient and inadequate for them to assess their knowledge or guide their further study. Creating effective questions is time-consuming because it may require considerable resources and skill in critical thinking (McComas and Abraham, 2005). The questions have to be carefully defined in order to accurately represent the intended learning outcome and the subject matter content involved.

COMBA is one of the applications of an adaptive assessment system. COMBA aims to provide a system which is able to accommodate complicated competencies, link competencies adequately, and support tracking of the knowledge state of the learner. The system focuses on the identification and integration of appropriate subject matter content (represented by a content taxonomy) and cognitive ability (represented by a capability taxonomy) into a hierarchy of competencies. The results of the system are the generated questions and tests. ASDEL is deployed as a stand-alone web application in order to deliver the tests to the learners. ASDEL is responsible for allowing a learner to view a question and to answer it. COMBA presents one possibility of using ontologies to automate question generation in adaptive assessment systems. It also offers interoperable, portable, and reusable resources for e-learning and knowledge management applications that define and update knowledge throughout a learner's life.

6.4.3 Knowledge engineering

In order to support lifelong learning, assessment systems have to focus on representation and updating a variety of knowledge domains, rules, assessments and learners' competency profiles. Representation of knowledge and competency is a crucial area in e-learning (Paquette, 2007). Without a good representation of the knowledge and competency to be processed, a delivery system will be unable to help its users according to their present and expected competency state. The competencies are expressed in the existing competency standards such as IMS RDCEO and HR-XML as simple, plain-language sentences. Without a structural model for a competency, the subject matter, cognitive ability, and target competencies will be unstructured textual fragments which are difficult to use in e-Learning systems, particularly in assessment systems. For example, consistency checking, assessing differences in knowledge levels, and comparing achievement in related domains remain essentially impractical.

The association between learning resources (documents, tools, actors, activities) and the knowledge and competencies they possess, contain, or process, is a key challenge that Semantic Web technologies can address. Drawing on the proposed competency model, an ontology is derived for competency modelling that combines the concepts of subject matter, cognitive ability, and other objects such as contexts, situations and tools. Such models provide ways to define competencies of individual learners, prerequisites and goals for resource content, the learner's knowledge state, and personalization capabilities for e-learning.

In this study, the knowledge domain was structured in as a "domain ontology". The domain subject matter content, capability taxonomy, and competence, were represented using SKOS. A domain expert expressed the domain content, the capability taxonomy, and competences in an English-like form. A knowledge engineer represented these elements in the form of a semantic network, and then transformed them into an ontology. The ontologies adhered to the criteria of ontology design: clarity, coherence, extendibility, minimal encoding bias, and minimal ontological commitment (Kalfoglou, 2001).

6.5 The improved model and the existing standards

How well the existing competency standard and the proposed competency model map the criteria of usability are analysed in Table 6-2 (Sitthisak *et al.*, 2007a), based on observation and implementation.

Criteria	IMS RDCEO	HR-XML	COMBA
Reusability : link to other competencies	Embed relation within competency record	Embed relation within competency record	Separate relation from competency record
Reusability : link to content	Embed subject matter content within itself	Refer to taxonomy	Refer to taxonomy
Interoperability: focus on	Interoperability definitions of competency	Interoperability measurement of competency	Interoperability definitions and measurement of competency
Equivalency and Similarity by focus on evidence	Unstructured definition on evidence element	Using Evidence element	Using Evidence element
Assessment request: measurability	Unstructured definition on weight, threshold and scale	Using competency weight element	Using scale and threshold of proficiency level
Assessment request: measurable behaviours	Unstructured definition	Depending on reference taxonomy	Using taxonomy of capability
Defining domain and scope of ontology	Depending on each system by using unstructured definition	Define structured definition by using competency description	Define structured definition by using capability, proficiency, situation, source, tools and subject matter content
Personalization : tracking knowledge state of learner	Depending on each system by using unstructured definition	Define structured definition by using competency description, weight and evidence	Define structured definition by using capability, proficiency, situation, tools and subject matter content

Table 6-2 A comparison of usability criteria in IMS RDCEO, HR-XML and COMBA

The criteria used in the table for evaluating the usability of IMS RDCEO, HR-XML and the proposed competency model are as follows.

1. Reusability

Reusability refers to linkages between competencies within a competency hierarchy and between competencies with content.

2. Interoperability – self-evident
3. Equivalency and Similarity by focus on evidence

The simplest possibility is for each competency to be associated with some information stating which similar definitions, held by other authorities, are accepted by the first definition owner as equivalent. If two systems are used with the intention of referring to the same competency concept, using different identifiers, a competency schema could allow automatic checking of whether either, or both, of the authorities maintain that the two concepts are equivalent.

4. Assessment request in terms of measurability and measurable behaviours

One of the objectives of developing a competency schema is to allow the capture of information about evidence used to rank, compare, and evaluate the sufficiency or desirability of a competency.

5. Defining domain and scope of ontology

The development of an ontology is suggested by defining its domain and scope, that is, by answering several basic competency questions (Noy and McGuinness, 2001). These questions will serve as an effective way of later proving: Does the ontology contain enough information to answer these types of questions? Do the answers require a particular level of detail or representation of a particular area?

6. Personalization in terms of tracking the knowledge state of the learner

One of the objectives of developing a competency schema is to define the concept of competence relating to a person's competencies and the proficiency level of that person with respect to a context (Cheetham and Chivers, 2006). The competency model should support defining the knowledge state of the learner in order to reflect a model of reality.

Semantic data transformation plays an important role in realizing the vision of the Semantic Web. It supports the transformation of data in different representations into ontologies (Bizer *et al.*, 2005). Many knowledge society services and computerised tools have been developed to help schools, universities, and organisations define and manage the competencies of their staff. These services and tools are looked upon as the main asset of an organisation from a knowledge management perspective. Different organisations or communities of practice may use different competency standards, vocabulary and processes for the representation of competency. Hence, a flexible competency ontology is needed in order to map different vocabulary into instances of different classes according to an organisation's particular perspective.

Table 6-3 shows the mapping of the categories in the RDCEO and HR-XML standards into the COMBA ontology. COMBA ontology classes fit well with these categories.

Competency Ontology Classes	**RDCEO Elements**	**HR-XML Elements**
Competence	Identifier, Title, Description	Name, Description, CompetencyId,
Subject Matter Content	Definition	TaxonomyId
Capability	Definition	TaxonomyId
Context	Metadata	Required, CompetencyWeight, CompetencyEvidence, UserArea
Component of competence	–	Competency

Table 6-3 Mapping RDCEO and HR-XML elements to competency ontology classes

6.6 Summary

The criteria for a good competency model are a rich data structure, flexibility, reflecting the reality of learner's competency, and involving specific, identifiable and measurable behaviours.

An improved competency model involves a capability associated with subject matter content, attitude, a proficiency level, evidence, any required tools, and definition of the situation which contextualises the competency. This competency model supports activity-based teaching and learning, the identification and integration of subject matter content within a broader teaching and learning context, and identification of the assessment.

In addition, this chapter mainly introduces the criteria for evaluating the usability of existing e-Learning competency standards (IMS RDCEO, HR-XML) and COMBA. These are reusability, interoperability, equivalency, measurability, domain and scope definitions of ontology, and tracking the knowledge state of the learner. These e-Learning competency standards are not able to accommodate complicated competencies and link competencies adequately. They also do not support comparisons of competency data between different communities, nor do they support tracking of the knowledge state of the learner. Therefore, the improvement of the competency model is proposed based on these current deficiencies. COMBA ontology classes also fit well with the categories of the RDCEO and HR-XML standards. The next chapter outlines and discusses implementation of the COMBA model, the system architecture and its components' functionality.

Chapter 7 COMBA system

7.1 Introduction

This chapter presents an overview of the COMBA system, the system architecture and its components' functionality. Next, a detailed description of data creation, representation and storage, are discussed, followed by the method of generating questions and standardising questions, adaptive sequence, and methods of question delivery.

7.2 System Overview

COMBA aims to provide a system which is able to accommodate complicated competencies, link competencies adequately, and support tracking of the knowledge state of the learner. The system focuses on the identification and integration of appropriate subject matter content (represented by a content taxonomy) and cognitive ability (represented by a capability taxonomy). This makes it easy to identify the assessment that would demonstrate successful teaching and learning.

The system was built on an ontological database that describes all resources and the relationships between them. The advantage of ontological schemas over database schemas is that the former define explicit formal specifications and include machine-interpretable definitions, to enable sharing common understanding of the structure of information among people or software agents. Thus, the ontological database is flexible and extensible, allowing the resources in the system to be described on the Semantic Web, interoperability between different systems, and reasoning about the described resources.

An assessment for a competency often actually tests component competencies. For example, a statistics course may test knowledge of the confidence interval[1] (Field, 2005) by testing the learners' ability to calculate, explain, and define the confidence interval in a variety of situations. A generic assessment item can be directly formulated from a competence specification by using the parameters of that competence: capability, subject matter content, and other elements such as the situation. For example, the assessment corresponding to the learning outcome, "learners understand the concept of a confidence interval" might be something like "Calculate the confidence interval for the following situation", or "Explain the importance of the confidence interval in the following situation", or "Define standard error".

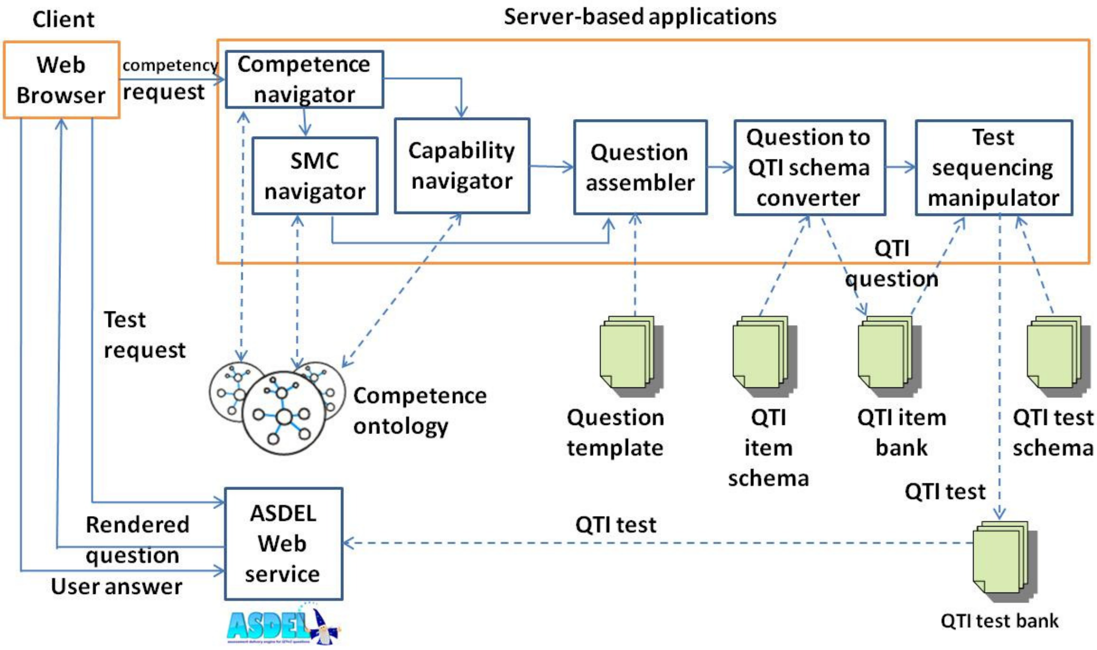


Figure 7-1 Architecture of the COMBA system

The COMBA implementation consists of a number of modules (Sitthisak, Gilbert and Davis, 2008a), illustrated in Figure 7-1. The Competence navigator is responsible for retrieving the requested competence, based on the domain request from the learner, and passing the competence to the Subject Matter Content and Capability navigator modules. The relevant subject matter and capability data received from those modules, together with the authoring question template files, are assembled to generate questions derived from the matrix of competencies crossed with cognitive abilities. Given a question which is now ready for further use, it is formatted using the QTI specification.

[1] A confidence interval is an interval estimate of a population parameter, and is used to indicate the reliability of an estimate.

The QTI specification facilitates the sharing of questions and tests, enabling investment in the development of common tools such as web-based authoring and delivery applications. For an adaptive test, this specification supports the use of pre-conditions and branching, allowing the embedding of sequencing and adaptive logic into a test. Adaptivity is limited to the questions referred to within the test. As a result, if the learner answered, it may not be possible to branch in directions not provided in the test. In addition, the inability to import external data may limit adaptivity.

In order to develop a test, the generated questions are linked together for storing in a test bank. For the delivery of the test, the system deploys an assessment delivery service (ASDEL) to allow a learner to view a question, to answer it, to receive feedback, and to view the assessment results.

7.3 Data creation, representation and storage

The domain subject matter content, capability taxonomy, and competence are based on the Simple Knowledge Organisation System (SKOS) (W3C, 2005). SKOS is used to express the basic structure of content, capability, and competence. Subject matter content is represented in the form of an ontology, based on the structure of its domain.

In general, a domain expert would express domain content, the capability taxonomy, and competence in an English-like form. A knowledge engineer would represent these elements in the form of a semantic network, and then transform them into an ontology. The ontologies adhere to the criteria of ontology design: clarity, coherence, extendibility, minimal encoding bias, and minimal ontological commitment (Kalfoglou, 2001). Sharing and reuse of information are integral aspects of the Semantic Web. In the COMBA system, the ontology was based on OWL-Lite (W3C, 2004a) which was sufficiently expressive to describe the subject matter hierarchy and provides for higher performance reasoning.

The framework of the ontologies is implemented in Protégé 3.3[2]. The Protégé tool supports knowledge acquisition and knowledge base development (Gennari *et al.*, 2003). Protégé includes an ontology editor and a system for generating and custom-tailoring forms for data entry by domain specialists. The ontology of the COMBA system is shown in Figure 7-2 (Sitthisak, Gilbert and Davis, 2008b). The definitions of

[2] http://protege.stanford.edu/

the elements in the competence ontology are shown in Table 7-1. An example fragment of the competency ontology declaration is shown as the OWL code in Figure 7-3.

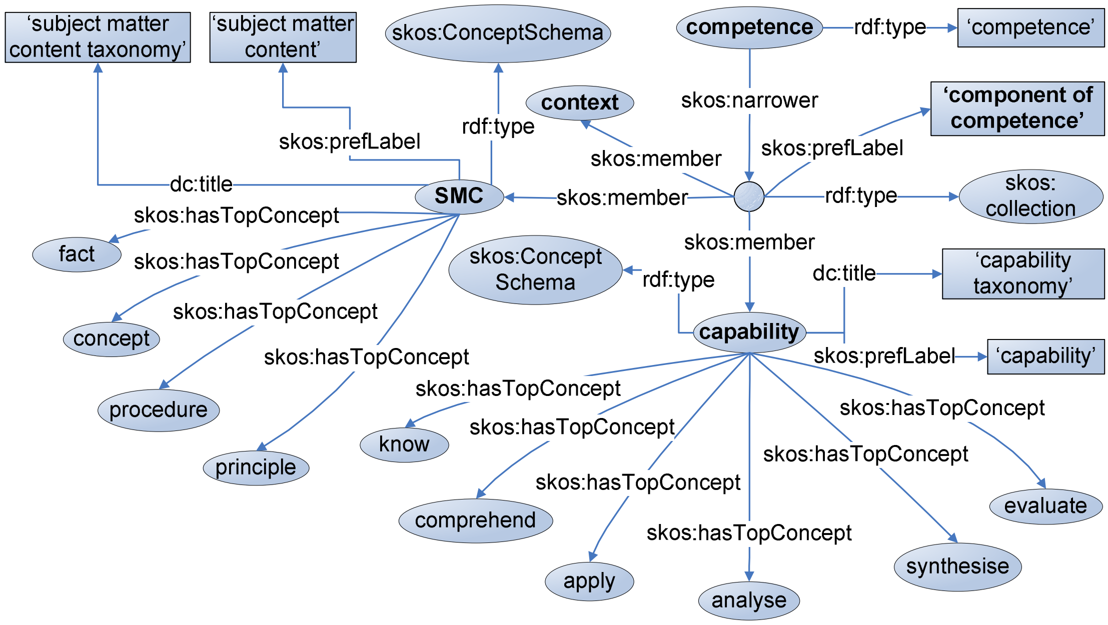


Figure 7-2 Ontology of COMBA

Element	Definition
Competence	Defines a capability associated with subject matter content, a proficiency level, evidence, any required tools, and definition of the situation which contextualises the competency.
SMC	Defines the subject domain of what the learner can do by the end of the unit of teaching and learning.
Capability	Defines behaviour that can be observed, based on a taxonomy of learning such as Bloom's, Gagné's nine areas of skill, or Merrill's cognitive domain.
Context	Defines the particular context and conditions of the competency, such as tools and situations.
Fact	Defines statements, or factual information, which consists of an attribute and a value.
Concept	Defines a group of objects or ideas which are designated by a single word or term. A concept has a number of attributes which are used to classify or categorise objects according to their values.
Procedure	Defines a sequential set of steps to accomplish a task or make a decision.
Principle	Defines cause-effect relationships describing the behaviour of a system. It can usually be expressed as some sort of an equation if the system is in the scientific or engineering domain.
Know, Comprehend, Apply, Analyse, Synthesise, and Evaluate	Cognitive domain capabilities according to Bloom.

Table 7-1 The definitions of each element in the competence ontology

```
<?xml version="1.0" ?>
- <rdf:RDF xmlns:p1="http://www.owl-ontologies.com/Comp.owl#rdfs:"
    xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:assert="http://www.owl-ontologies.com/assert.owl#"
    xmlns="http://www.owl-ontologies.com/Comp.owl#"
    xml:base="http://www.owl-ontologies.com/Comp.owl">
  + <owl:Ontology rdf:about="">
  + <owl:Class rdf:ID="Procedure">
  + <owl:Class rdf:ID="Apply">
  + <owl:Class rdf:about="#Synthesise">
  + <owl:Class rdf:ID="Relation_Att_Value">
  + <owl:Class rdf:about="#Fact">
  + <owl:Class rdf:about="#SMC_taxonomy">
  + <owl:Class rdf:about="#Comprehend">
  + <owl:Class rdf:about="#Capability_taxonomy">
  + <owl:Class rdf:about="#Evaluate">
  + <owl:Class rdf:about="#Step">
  + <owl:Class rdf:about="#Analyse">
  + <owl:Class rdf:about="#Concept">
  + <owl:Class rdf:about="#Competence">
  + <owl:Class rdf:about="#RelationWithOther">
  + <owl:Class rdf:about="#Group_step">
  + <owl:Class rdf:about="#Principle">
  + <owl:Class rdf:about="#Know">
  + <owl:ObjectProperty rdf:ID="a_sub_issue">
  + <owl:ObjectProperty rdf:about="#SMC">
  + <owl:ObjectProperty rdf:ID="Sub_issue">
```

Figure 7-3 Sample fragment of OWL code for competency ontology

Protégé stores OWL ontologies in tables. Eight tables were implemented in the COMBA system, including three tables for the capability ontology, four tables for the categories (fact, concept, procedure, principle) of the subject matter ontology, and one table for the competency ontology.

Three tables of the capability ontology were implemented as ‘capability category’, ‘capability key verbs’, and ‘capability ordering’. The ‘capability category’ table referred to the six capability categories in the capability taxonomy. These were ‘know’, ‘comprehend’, ‘apply’, ‘analyse’, ‘synthesise’, and ‘evaluate’. The ‘capability key verbs’ table referred to the key verbs in each capability category such as ‘explain’, ‘calculate’, and ‘define’. The ‘capability ordering’ table linked two capabilities such that the first capability must be mastered before the next one. For example, the ‘comprehend’ capability must be mastered before the ‘apply’ capability.

The ontology repositories for the COMBA system were native stores. Native stores are directly built on the file system, thereby contributing positively to the load reduction and update time (Atanas, Damyan and ManovDimitar, 2005). In order to populate the OWL models, store them in a native store, and query them with a program, the Jena Semantic Web Framework[3] was implemented. Jena is an inference engine which uses the SPARQL[4] query language to extract data from the ontology. SPARQL follows SQL-style syntax such as using SELECT to process the learner query. An example is shown in Table 7-2.

```
SELECT ?capability ?SMC ?imply ?context ?relatedSMC
WHERE {?competence Comp:SMC ?SMC
            ?competence Comp:capability ?capability
  ?competence Comp:Context ?context
  ?capability Comp:imply ?imply
{?SMC Comp:the_input ?relatedSMC} UNION
{?SMC Comp:RelatedConcept ?relatedSMC}};
```

Table 7-2 An example of the SPARQL query language

In this research, the ‘SELECT’ SPARQL query was used to extract data from RDF/OWL code, returning it as a tabular result set. The ‘UNION’ SPARQL query was used to combine result sets into a larger result set. SPARQL result sets were serialised into XML format to allow their direct manipulation. The ‘SELECT’ part declares the variables to be output by the query, in this case, the variables named ‘capability’, ‘SMC’, ‘imply’, ‘context’, and ‘relatedSMC’. The query finds statements in the graph for which all of the triples in the “WHERE” clause hold. In this query, six triple patterns were included. An important point is that within a collection of the triple patterns, a variable must have the same value no matter where it is used. So, the variable ‘competence’, ‘capability’ and ‘SMC’ would always be bound to the same resource. In other words, this query will match any resource that has all of the desired properties. A resource that does not contain all of these properties will not be included in the results because it won’t satisfy all of the triple patterns. However, the ‘UNION’ query allows optional matching. In this example, the ‘UNION’ keyword joined two query patterns. If an element resource matches either of these patterns, then it will be included in the query solution.

[3] http://jena.sourceforge.net/

[4] http://www.w3.org/TR/rdf-sparql-query/

The result of a SPARQL 'SELECT' query is a sequence of results that, conceptually, form a table or result set. Each row in the table corresponds to one query solution. And each column corresponds to a variable declared in the 'SELECT' clause. Figure 7-4 shows a fragment of a SPARQL query result set in XML format.

```
  <?xml version="1.0" ?>
- <sparql xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xs="http://www.w3.org/2001/XMLSchema#"
    xmlns="http://www.w3.org/2005/sparql-results#">
  - <head>
      <variable name="competence" />
      <variable name="capability" />
      <variable name="SMC" />
      <variable name="imply" />
      <variable name="context" />
    </head>
  - <results>
    - <result>
      - <binding name="competence">
          <uri>http://www.owl-
            ontologies.com/Comp.owl#calculate_margin_of_error</uri>
        </binding>
      - <binding name="capability">
          <uri>http://www.owl-
            ontologies.com/Comp.owl#Calculate</uri>
        </binding>
      - <binding name="SMC">
          <uri>http://www.owl-
            ontologies.com/Comp.owl#a_confidence_interval</uri>
        </binding>
      - <binding name="context">
          <literal xml:lang="en">Nine hundred (900) high school
            freshmen were randomly selected for a national survey
            Among survey participants the mean grade-point average
            (GPA) was 2.7, and the standard deviation was 0.4
            assuming a 95% confidence level.</literal>
        </binding>
      </result>
```

Figure 7-4 Sample fragment of a SPARQL query result set in XML format

7.4 Method of generating questions

In any unit of teaching and learning, there are usually a number of competencies that it is intended learners achieve. These competencies and their linkages may be assembled into trees. While the relationship between competence nodes may be modelled as a family relation such as parent and child, there is no necessary ordering of the nodes on the same level, thus yielding a tree structure rather than a hierarchy. Given competencies assembled as a tree structure, it is assumed that proficiency in all children of a defined competency is a necessary precondition for achieving proficiency in the parent. While the tree structure defines a structure which may be traversed top-down, bottom-up, leftwards, or rightwards, it does not imply sequencing and does not imply a starting point for questions. In a top-down approach, the competency tree might be used to drill down into the component competencies for a target competency, helping to define the details of what to test for that high-level competency. Alternatively, in a bottom-up approach, the competency tree might be traversed from a leaf node up to parent nodes. This may help to define a test for a range of higher-level competencies given a lower-level competency. To generate a particular series of the questions, a particular traversal algorithm and a starting point must be selected and implemented.

Competencies based on COMBA could be implemented for any domain, where these can be expressed in the form of 'subject matter' and 'capability' taxonomies, for example, these are shown in Table 7-3. Figure 7-5 represents these competencies graphically.

Competence	**Subject Matter**	**Capability**	**Context**	**Sub-competence**
Students can calculate the confidence interval	Concept: the confidence interval	Apply: Calculate	Nine hundred high school first year students were randomly selected for a national survey. Among survey participants, the mean grade-point average was 2.7, and the population standard deviation was 0.4 assuming a 95% confidence level.	Students can calculate the standard error
Students can calculate the standard error	Concept: the standard error	Apply: Calculate	(same as above)	–

Table 7-3 Examples of confidence interval competency represented in the competency model

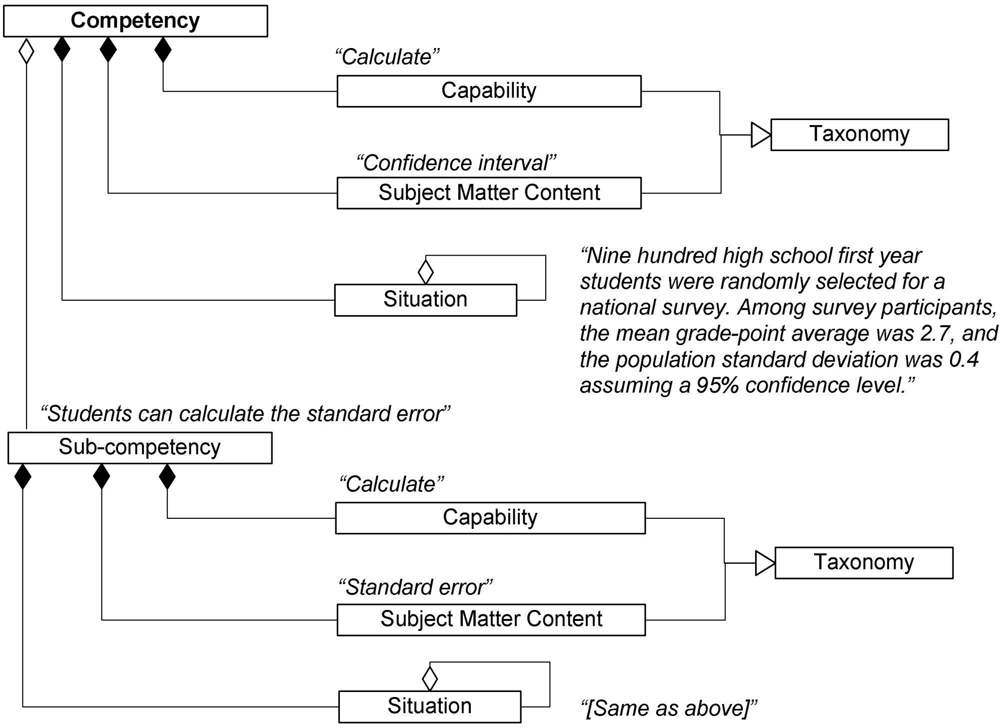


Figure 7-5 Example of competency

Question generation begins from the competency of interest, submitted to the system as shown in Table 7-3. The Competence Navigator module (shown in Figure 7-1) retrieves subject matter as shown in Table 7-4 and capability nodes relevant to the competency as shown in Table 7-5, using the competency ontological database, as discussed in section 7.3. Figure 7-6 represents the subject matter for Table 7-4 graphically. Figure 7-7 represents the capability for Table 7-5 graphically.

Subject Matter	**Related subject matter**
Concept: the confidence interval	Concept: the standard error Fact: the alpha value Fact: the critical z score
Concept: the standard error	Fact: the measure of dispersion Fact: the sample size

Table 7-4 Some examples of subject matter content based on the confidence interval topic

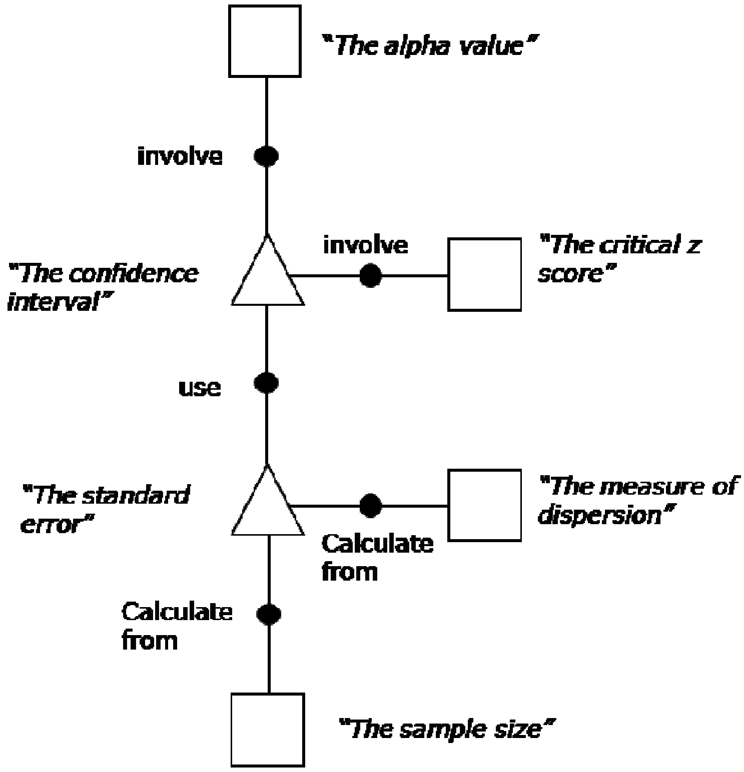


Figure 7-6 Example of the related topics of the confidence interval topic

Capability	Supporting capability
Apply: Calculate	Comprehend: Explain
Comprehend: Explain	Know: Define

Table 7-5 Some examples of capabilities based on the confidence interval topic

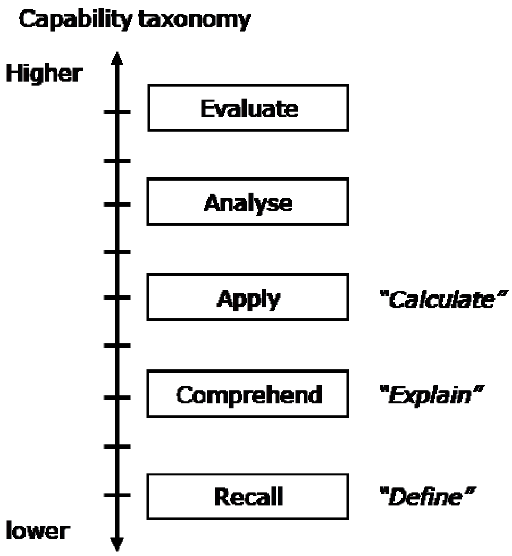


Figure 7-7 Example of the related capabilities of the competency

Given the subject matter and capability of the submitted competency, the related topics in the four subject matter category tables and capability in the 'capability ordering' table are retrieved as well. For example, if the requested subject matter is 'confidence intervals', the retrieved related subject matter includes 'critical z score' and 'standard error'. For the 'calculate' capability, 'explain' and 'define' capabilities were retrieved as well.

Question templates are used to assemble the retrieved subject matter and capability nodes into questions as shown in Table 7-6 (Sitthisak, Gilbert and Davis, 2009). For example, given the 'confidence interval' competency, the related subject matter and capabilities are inserted into the question templates to yield questions such as 'Explain the importance of the critical z score', as shown in Figure 7-8, and examples of generated questions as shown in Table 7-7.

Template No.	Question Templates
1	[Capability] + [Subject Matter]
2	[Capability] + [Related Subject Matter]
3	[Capability] + [Subject Matter] + [Situation]
4	[Capability] + [Related Subject Matter] + [Situation]

Table 7-6 Question templates

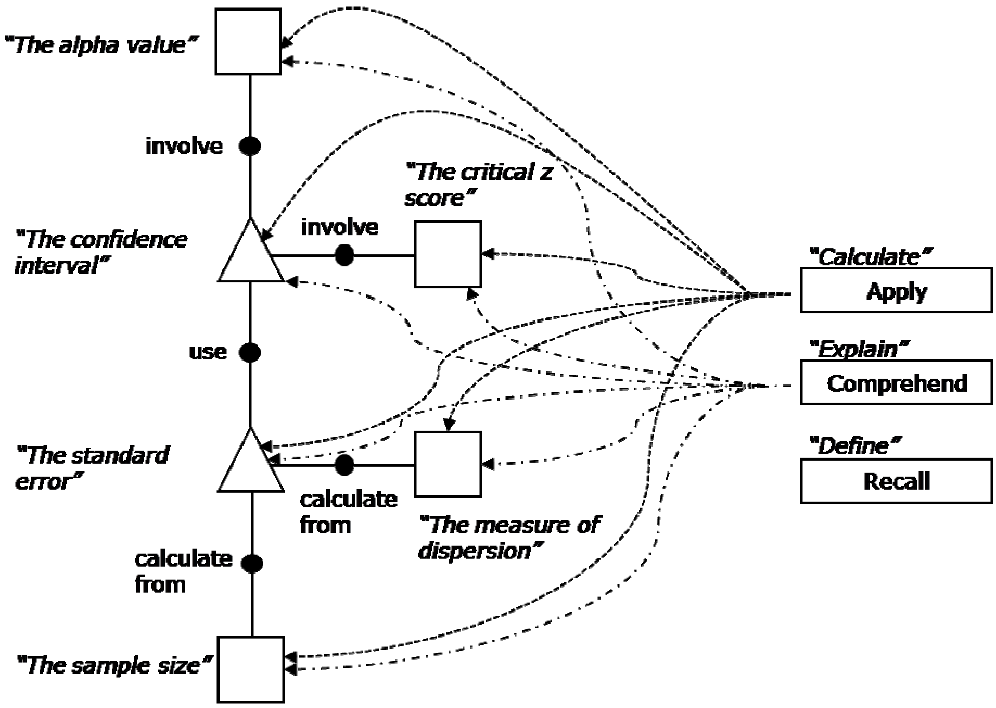


Figure 7-8 Example of assembling the retrieved subject matter and capability nodes into questions using template no.1 and template no.2

Competence	Generated question	Question Templates No.
Students can calculate the confidence interval	Calculate the confidence interval Calculate the critical z score Calculate the alpha value	1
	Calculate the standard error Calculate the measure of dispersion Calculate the sample size	2
	Explain the importance of the confidence interval Explain the importance of the critical z score Explain the importance of the alpha value	1
	Explain the importance of the standard error Explain the importance of the measure of dispersion Explain the importance of the sample size	2

Table 7-7 Sample generated questions

The process of traversing competencies, retrieving the relevant nodes, and converting these to questions is recursive. The generated questions are transformed for conformance to the QTI specification by a conversion process using the QTI schema.

The relatively unsophisticated method of generating questions, in particular the use of simple question templates, yields some questions which are inappropriate, do not make good sense, or show poor grammar and syntax such as ‘Calculate ER Diagram’. The generated questions needed to be filtered by a domain expert before the system automatically constructed a test.

7.5 Adaptive sequence

Within a test constructed according to the IMS QTI specification, the sequencing and adaptive logic are expressed in branching rules. For example, an adaptive sequence may provide a question at a slightly higher level if a learner succeeds or a question at a lower level otherwise. The starting point for questioning may be at a target competency level or a component of a target competency level. In this example, the starting question may be at particular level, such as question1 (shown in Figure 7-9) in the portion labelled A.

```
  <?xml version="1.0" encoding="UTF-8" ?>
- <assessmentTest xmlns="http://www.imsglobal.org/xsd/imsqti_v2p1"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.imsglobal.org/xsd/imsqti_v2p1
    http://www.imsglobal.org/xsd/imsqti_v2p1.xsd" identifier="TEST"
    title="Function Point Analysis Test">
  - <testPart identifier="part1" navigationMode="linear" submissionMode="individual">
      <itemSessionControl showFeedback="true" />
    - <assessmentSection identifier="sectionquestion1" title="Section question:"
        visible="true">
        <assessmentItemRef identifier="question1" href="question1.xml" />          A
      </assessmentSection>
    - <assessmentSection identifier="sectionRquestion1" title="SectionR question1"
        visible="true">
      - <branchRule target="EXIT_TEST">
        - <equal toleranceMode="exact">
            <variable identifier="question1.SCORE" />                              B
            <baseValue baseType="float">1.0</baseValue>
          </equal>
        </branchRule>
      </assessmentSection>
    - <assessmentSection identifier="sectionquestion2" title="Section question:"
        visible="true">
        <assessmentItemRef identifier="question2" href="question2.xml" />          C
      </assessmentSection>
    </testPart>
  + <outcomeProcessing>
  + <testFeedback access="atEnd" showHide="hide"
      outcomeIdentifier="outcomeIdentifier" identifier="outcomeValue" title="Test
      Feedback">
  </assessmentTest>
```

Figure 7-9 Example of QTI branching rules in XML format

Figure 7-9 presents the question file which is incorporated into the test by reference not by direct aggregation. Portions labelled A and C show the learner items called “question1” and “question2” respectively. The portion labelled B illustrates the branching rule.

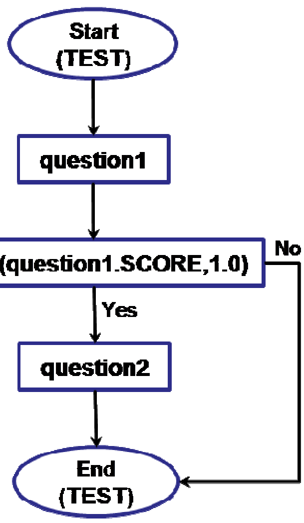


Figure 7-10 Flow of questions in a QTI test

For example, if the learner succeeds on question1, the test jumps forward to the end of the test (shown as branchRule target= ‘EXIT_TEST’) or goes to “question2” in the section labelled C otherwise. The ASDEL validator graph of this adaptive sequencing is shown in Figure 7-10.

In this research, the QTI questions are sequenced according to traversal algorithm and starting point.

The traversal algorithm specifies the method of selecting the next question to ask the learner, depending on their previous answer. For one traversal algorithm, if the learner succeeded on a question, a slightly more challenging question was presented next. This was a question at the same capability level and at a higher subject matter level than the previous question. If the learner failed the question, the system presented an easier question. This was a question at the same capability level and at the same or a lower subject matter level than the previous question.

For another algorithm, if the learner succeeded on a question, a slightly more challenging question was presented next. This was a question at the same or a higher capability level and at the same or a higher subject matter level than the previous question. If the learner failed the question, the system presented an easier question as in the first algorithm. This was a question at the same or a lower capability level but at a lower subject matter level than the previous question.

One starting point presented the first question from the lowest subject matter level and the lowest ability level. Another starting point presented the first question from the highest subject matter level and the highest ability level.

7.6 Method of question delivery

In this research, ASDEL was deployed as a stand-alone web application to deliver the tests to the learners. ASDEL allows a learner to view a question, to answer it, to receive feedback (shown in Figure 7-11; the xml for this question is in appendix A), and to view the test result (shown in Figure 7-12).

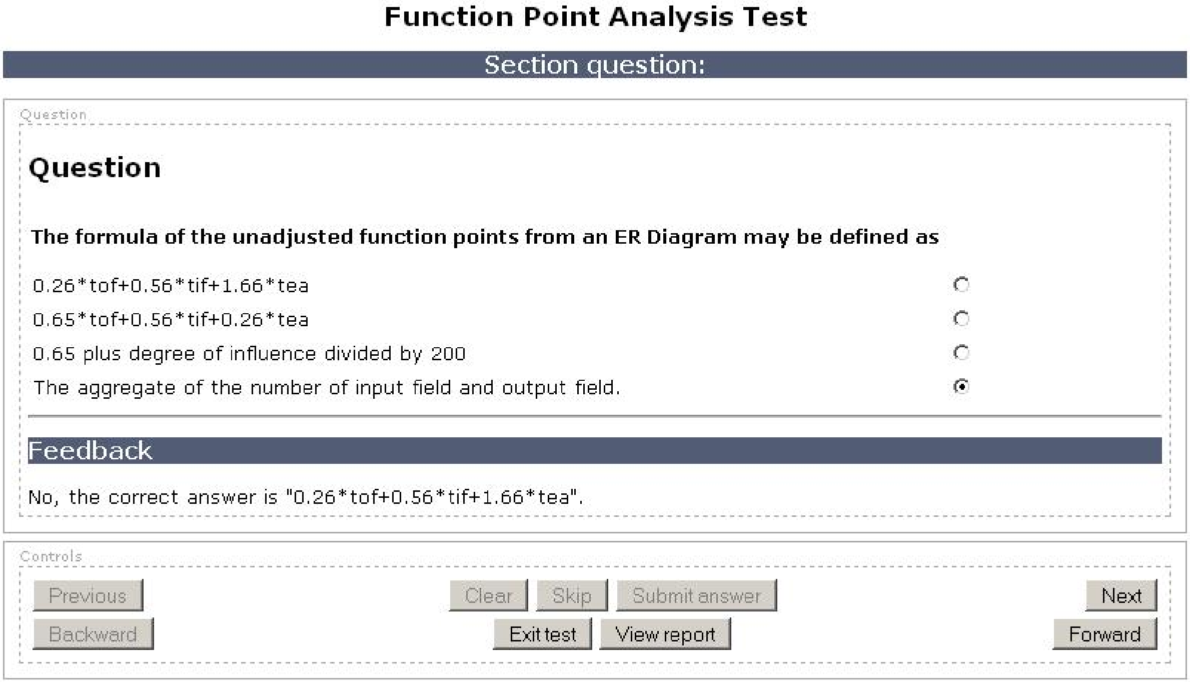


Figure 7-11 ASDEL displaying a question, receiving an answer, and giving feedback

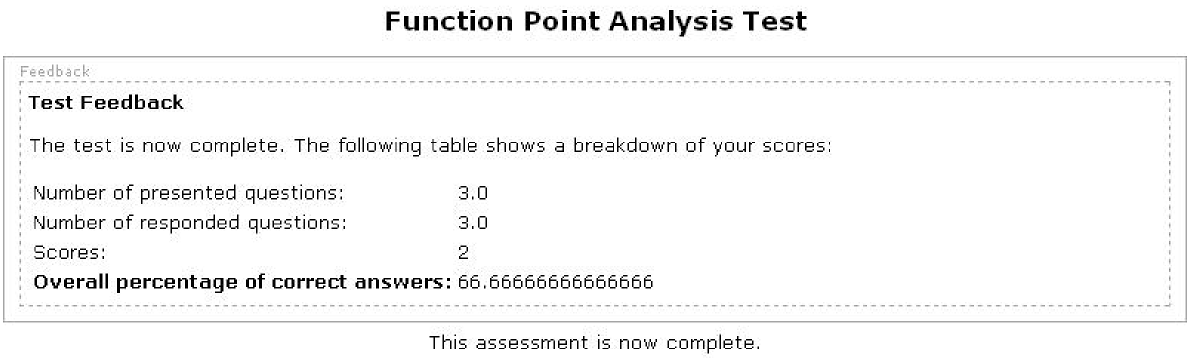


Figure 7-12 ASDEL showing the test result

7.7 Summary

This chapter describes the process of generating questions and tests within the COMBA system. The adaptive sequencing of the QTI test is expressed in pre-branching rules of QTI schema. While the system successfully demonstrates a data model and method of automatically generating questions, the immediate challenge of dealing with the generating mechanism is raised. The generated questions must ensure standards of English grammar and syntax. Three experiments demonstrating the generation of tests, the sequencing of generated test from a competency data model, and comparing a variety of the adaptive sequences, are outlined in the following chapter.

Chapter 8 COMBA system experiments

8.1 Introduction

This chapter outlines the three experiments that were carried out to demonstrate the generation of assessments, the sequencing of generated assessments from a competency data model, and to compare a variety of the adaptive sequences. For each experiment, the methods and experimental results are described.

8.2 Research design

All experiments involved students from the School of Electronics and Computer Science at the University of Southampton. Ethical approval was sought and obtained from the Ethical Committee of the School for the students to participate in the study.

For each experiment, students were given the opportunity at the end of their lecture to use a prototype to help them identify, diagnose and understand the boundaries of their own competencies.

A questionnaire (see appendix B, C, D, and F) given to each student explained the objectives and relevance of the study, assured the students of anonymity, gave them the option of not participating in the study if they did not wish to, and asked them to evaluate their experiences in using the system. No information was asked which could be used to identify the participant. The data from the questionnaires was analysed using SPSS.

To calculate an expected sample size for the three experiments, G*Power was used (Faul *et al.*, 2007). G*Power is a general power analysis program which computes sample sizes for given effect sizes, alpha levels, and power values. The effect size expresses whether the difference observed is a difference that matters. The larger the effect size, the easier it is to see that there is a difference between the two means being compared. Effect sizes from 1 to 2 are typical for exploratory study. For this study, effect size was set to 1.5. The value of α was 0.05, and the required power was 0.8, with a two-tailed test. The expected sample size was calculated as n = 18 by the program. This means at least 18 students are needed for each experiment to detect effect sizes of 1.5 with 80% power.

8.3 Aims of the experiments

The aims of the experiments were as follows.

1. To determine the opinions of students about the quality of the generated questions.
2. To determine the opinions of students about the usefulness and adaptiveness of the COMBA generated test.
3. To determine the opinion of students about the adaptive sequencing of the COMBA generated test.
4. To explore the relationship between a variety of adaptive sequences and their pedagogical effectiveness.

8.4 Experiments and results

Table 8-1 provides a summary of the three experiments and their results. More details are described in the following sections.

<table>
<tr><th>EXPT</th><th>Aims</th><th>Experiment design</th><th>Results</th></tr>
<tr><td>1</td><td>To determine the opinions of students on the quality of the generated questions.</td><td>• The independent variable: the generated questions were either ‘generic’ or ‘specific’.
• The dependent variable: the student ratings of clarity, usefulness, challenge, and match with learning outcomes.</td><td>• The students rated the clarity of generic questions significantly higher than that of specific questions
• The students rated the challenge of the specific questions significantly higher than that of the generic questions.</td></tr>
<tr><td>2</td><td>To determine the opinions of students on the efficiency and effectiveness of the generated tests.</td><td>• The independent variables: student opinion compared against a “population” average.
• The dependent variables: the student ratings on twelve criteria of efficiency and effectiveness.</td><td>• The mean ratings of efficiency and effectiveness were significantly higher than the “population” average of 2.5 for nine of the twelve measured variables.</td></tr>
<tr><td>3.1</td><td>To determine the opinions of the <u>students </u>on two adaptive sequencing algorithms and two starting points for questioning.</td><td rowspan="2">• The independent variables:
o the algorithms for the next question were either ‘SameAbilityLevel’ or ‘HigherAbilityLevel’,
o the starting points were either ‘Top-down’ or ‘Bottom-up’, and
o control variable: the order of presentation of the adaptive system was starting from ‘Top-down’ for half the group, or from ‘Bottom-up’ for the other half.
• The dependent variable: the student ratings on the six criteria.</td><td>• Significant interaction between starting point order of presentation with a Top-down starting point presented first, mean percentage correct was lower than when this was presented second with a Bottom-up starting point presented first, mean percentage correct was higher than when this was presented second.
• Significant order of presentation effect: mean usefulness rating higher when the Top-down method was presented first than when presented second.</td></tr>
<tr><td>3.2</td><td>To determine the opinions of the <u>experts</u> on two adaptive sequencing algorithms and two starting points for questioning.</td><td>• the mean IdentKw, HelpLO, HelpTopic, CompelteAss, and UsefulSelf for ‘LowerAbilityLevel’ algorithm was significantly higher than for the ‘SameAbilityLevel’ algorithm.</td></tr>
</table>

Table 8-1 Summary of the three experiments and their results

8.4.1 Experiment 1

The first experiment was designed to demonstrate the generation of assessments from a competency data model and to explore the following questions:

- How well were the generated questions rated using the criteria of clarity, usefulness, challenge, and match with the learning outcomes?
- Were the generated questions semantically intelligible to an expert teacher of the domain?

The independent variable was the type of the generated question: ‘generic’ and ‘specific’. The dependent variables were the student ratings on the criteria of clarity, usefulness, challenge, and match with the learning outcomes.

The competencies were collected from the INFO1013 Tools and Techniques for IT Modeling course. Competencies based on COMBA could be implemented for any domain, where these can be expressed in the form of ‘subject matter’ and ‘capability’ taxonomies. The topics of the course involved confidence intervals and associated issues as following:

- critical z score
- Alpha value
- standard error
- measure of dispersion
- sample size

Subject matter content for the competency data model was collected from the core textbook and the website of the course syllabus, and reviewed by a domain expert in this field. Table 8-2 and Table 8-3 represent this data.

Competence	Students can calculate the confidence interval
Subject Matter	Concept: the confidence interval
Related subject matter	Concept: the standard error Fact: the alpha value Fact: the critical z score
Capability	Apply: Calculate
Supporting capability	Comprehend: Explain Know: Define
Context	Nine hundred high school first year students were randomly selected for a national survey. Among survey participants, the mean grade-point average was 2.7, and the population standard deviation was 0.4 assuming a 95% confidence level.
Sub-competence	Students can calculate the standard error

Table 8-2 Examples of confidence interval competency represented in the competency model

Competence	Students can calculate the standard error
Subject Matter	Concept: the standard error
Related subject matter	Fact: the measure of dispersion Fact: the sample size
Capability	Apply: Calculate
Supporting capability	Comprehend: Explain Know: Define
Context	Nine hundred high school first year students were randomly selected for a national survey. Among survey participants, the mean grade-point average was 2.7, and the population standard deviation was 0.4 assuming a 95% confidence level.
Sub-competence	-

Table 8-3 Examples of a standard error competency represented in the competency model

The system generated 42 questions within the confidence interval topic. These questions were reduced to 25 questions, based on a review by two domain experts and selection of the questions which would most appropriately address the experimental questions. Examples of the questions filtered out were 'Calculate the confidence interval', 'Calculate the critical z score' without a situation, and 'Define the meaning of the critical z score in this situation'. The review ensured a high standard of English grammar so that the questions were phrased precisely, clearly and were easy to understand (McComas and Abraham, 2005).

In this system, the question templates were used to assemble the subject matter and capability into questions as shown in Table 8-4, and examples of generated questions are shown in Table 8-5. The 25 questions were classified according to their type, whether they were 'generic' questions or 'specific' questions (see Table 8-4). There were 15 specific questions and 10 generic questions. There are some questions that the experts would have expected such as "What is the effect of sample size on the width of a confidence interval?" and "In computing a confidence interval, when do you use 't' and when do you use 'z'?". The topic found in these questions is not directly represented in the intended learning outcome and the subject matter involved. These may be called meta-questions. The use of the question templates did not allow the generation of such questions.

The questions involved three distinct capabilities: Define, Explain, and Calculate. The three 'specific' questions involved one question for each capability, while the two 'general' questions involved one question for each of the Define and Explain capabilities.

Question Templates	Type of Question
[Capability] + [Subject Matter]	Generic Question
[Capability] + [Related Subject Matter]	Generic Question
[Capability] + [Subject Matter] + [Situation]	Specific Question
[Capability] + [Related Subject Matter] + [Situation]	Specific Question

Table 8-4 Question templates

Learning outcome	Generated question	Type of Question
Students understand the concept of a confidence interval, and can calculate it. [from INFO1013]	Define the meaning of the confidence interval.	Generic Question
	Explain the importance of the critical z score.	Generic Question
	Calculate the confidence interval for this situation: Nine hundred high school first year students were randomly selected […] and the population standard deviation was 0.4 assuming a 95% confidence level.	Specific Question
	Explain the importance of the standard error in this situation: Nine hundred high school first year students were randomly selected […] and the population standard deviation was 0.4 assuming a 95% confidence level.	Specific Question

Table 8-5 Sample generated questions

The questionnaire (see appendix B) asked the students to rate five generated questions against the four criteria on a 3-point Likert scale ('Yes', 'No opinion', and 'No', coded as 1, 2, and 3 respectively). Each questionnaire comprised three specific questions and two generic questions. The 25 questions were distributed between five question papers (questionnaires) in order to reduce the workload and time taken of the students answering the questionnaire.

The participants were voluntary 1st year undergraduate students of the INFO1013 Tools and Techniques for IT Modeling course. The questionnaires were randomly distributed to all attending students at the end of a lecture. The study gathered data from 27 students. (Thirty students were enrolled and expected, but on the day three students failed to attend.)

SPSS generates reports for four test statistics for the multivariate test of differences in mean ratings of questions: Wilks's Lambda, Hotelling's Trace, the Pillai-Bartlett trace, and Roy's largest root. In this experiment, Wilks's Lambda and Hotelling's Trace are the best for our purpose because group differences are concentrated on the variate of rating classification (Field, 2005).

Effect	The statistic method	Value	F	Hypoth df	Error df	Sig.
Question type	Wilks's Lambda	0.888	4.023	4	127	0.004
	Hotelling's Trace	0.127	4.023	4	127	0.004
Capability type	Wilks's Lambda	0.940	0.992	8	254	0.443
	Hotelling's Trace	0.063	0.993	8	252	0.442
Question type * Capability type (interaction)	Wilks's Lambda	0.996	0.134	4	127	0.970
	Hotelling's Trace	0.004	0.134	4	127	0.970

Table 8-6 Multivariate Test

As can be seen in Table 8-6, the multivariate tests for differences in rating according to question type, capability type, and the question by capability type interaction, showed significance only for differences between question types (Wilks's Lambda p = 0.004 and Hotelling's Trace p = 0.004). Table 8-7 provides the estimated marginal means for the four ratings according to question type.

Dependent Variable	Question type	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Clear	Specific Q	1.975	0.071	1.834	2.116
	Generic Q	1.630	0.087	1.457	1.802
Useful	Specific Q	1.630	0.071	1.490	1.769
	Generic Q	1.759	0.086	1.588	1.930
Match to learning outcomes	Specific Q	1.877	0.070	1.738	2.015
	Generic Q	1.778	0.086	1.608	1.948
Challenging	Specific Q	1.346	0.057	1.233	1.459
	Generic Q	1.500	0.070	1.361	1.639

Table 8-7 Estimated Marginal Means for Question Type

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Question type	Clear	4.48	1	4.48	10.87	0.001
	Useful	0.33	1	0.33	0.83	0.37
	Match to learning outcomes	0.15	1	0.15	0.37	0.54
	Challenging	1.33	1	1.33	5.03	0.03
Error	Clear	53.56	130	0.41		
	Useful	52.52	130	0.40		
	Match to learning outcomes	51.93	130	0.40		
	Challenging	34.44	130	0.27		

Table 8-8 Tests of Between-Subjects Effects

Table 8-8 provides the tests of between-subject effects for question type, where it may be seen that there were significant differences in mean ratings of 'Clear' and 'Challenging', but there were no significant differences in mean ratings of 'Useful' and 'Match to learning outcomes'. Six out of eight of the 95% confidence intervals were below 2, indicating a tendency to rate "Yes" rather than "No opinion" or worse.

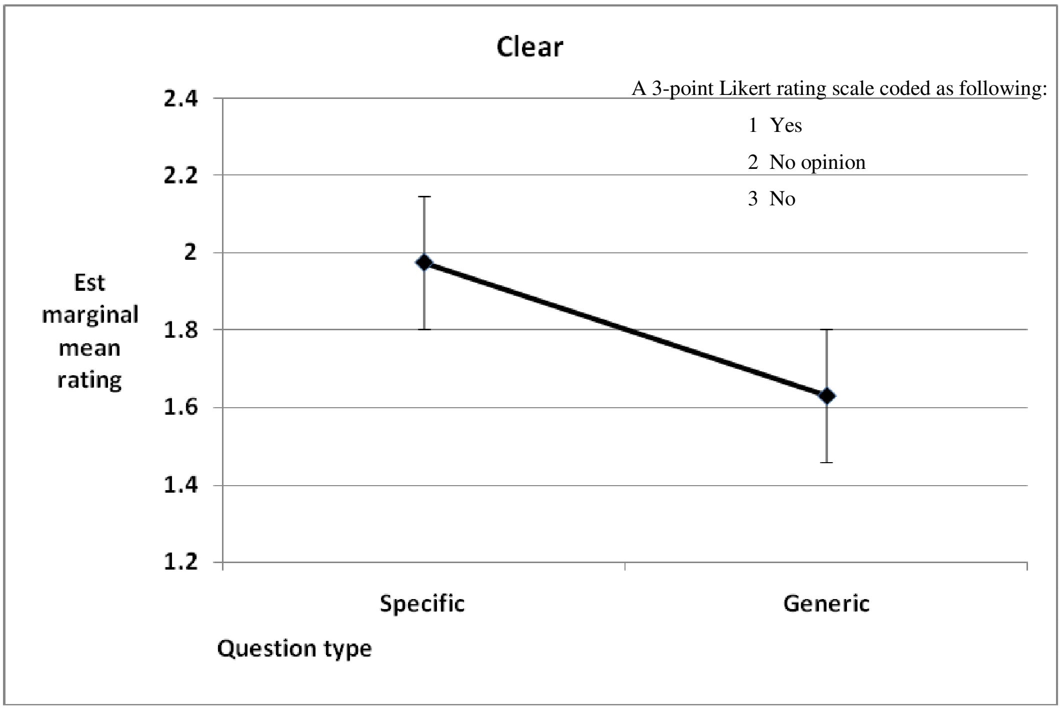


Figure 8-1 Means profile plots of estimated marginal mean Clear for question types

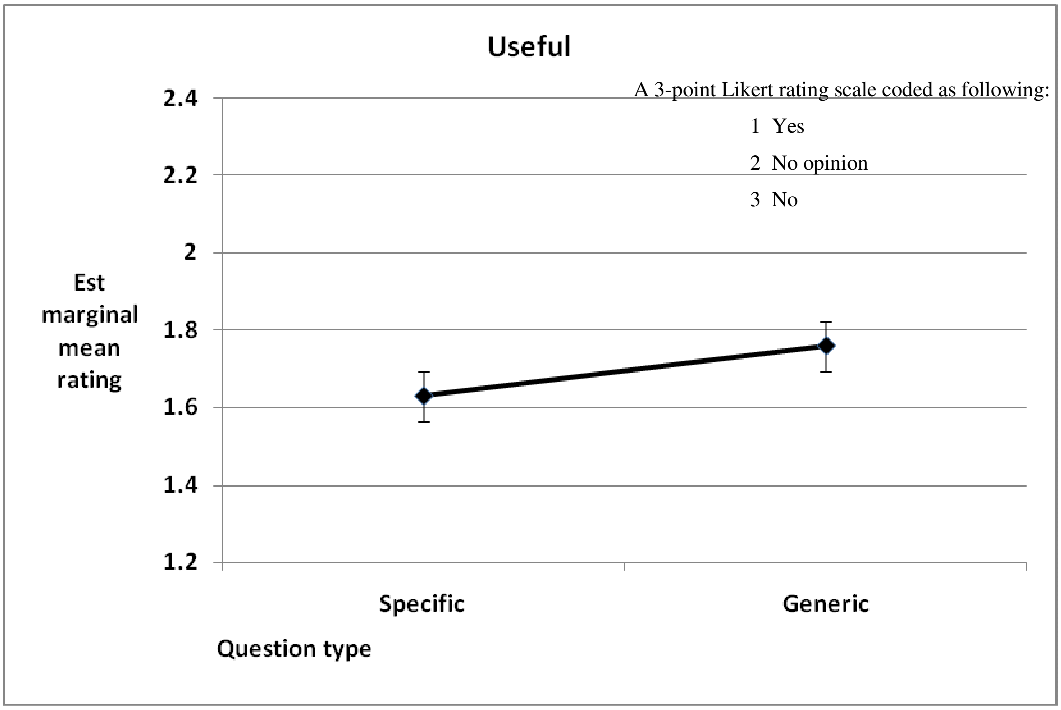


Figure 8-2 Means profile plots of estimated marginal mean Useful for question types

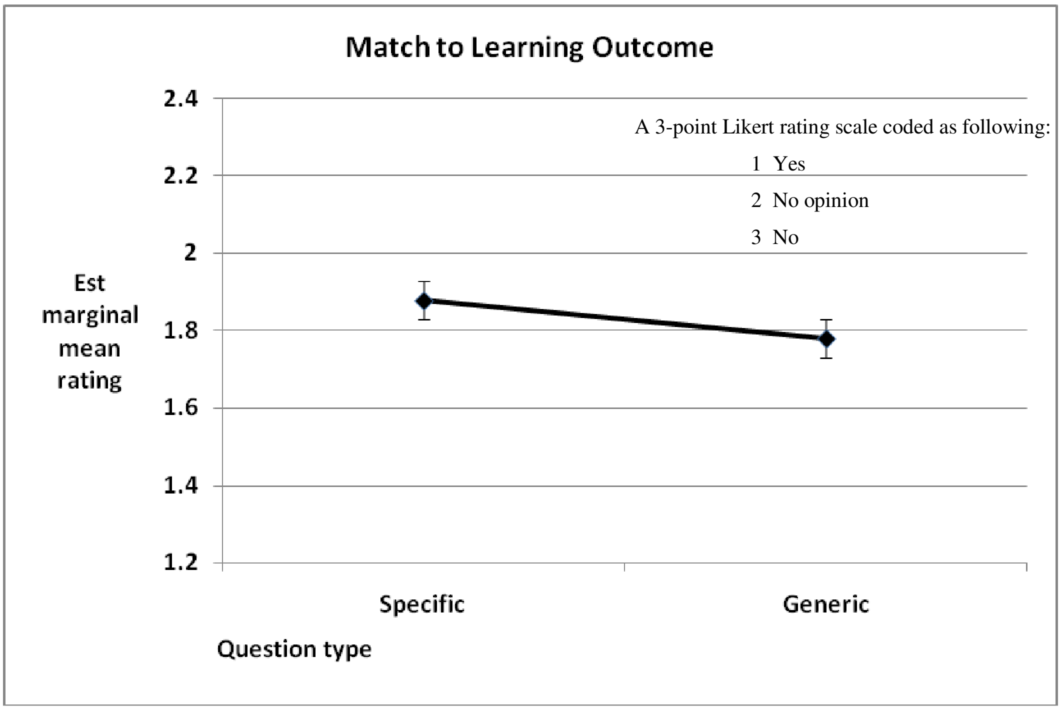


Figure 8-3 Means profile plots of estimated marginal mean Match to Learning Outcome for question types

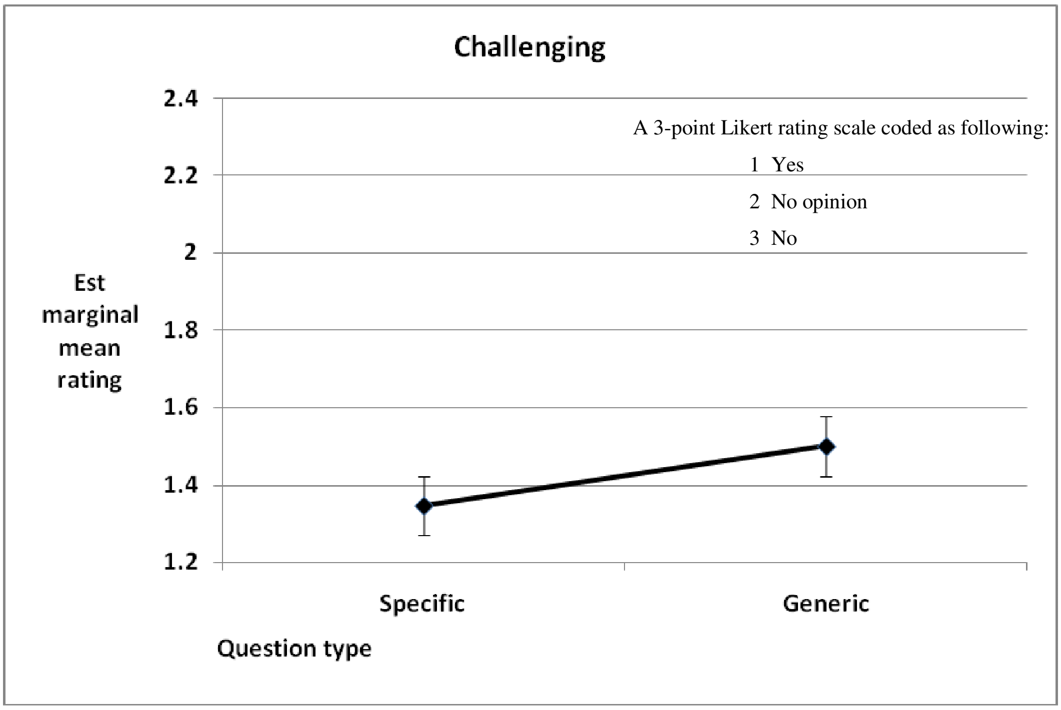


Figure 8-4 Means profile plots of estimated marginal mean Challenging for question types

An inspection of the profile graphs in Figure 8-1, Figure 8-2, Figure 8-3, and Figure 8-4, shows that the students rated the clarity of generic questions significantly higher than that of specific questions, while rating the challenge of the specific questions significantly higher than that of the generic questions. The students rated the specific and the generic questions as not significantly different with regard to mean ratings of 'Useful' and 'Match to learning outcome'.

8.4.2 Experiment 2

The second experiment was designed to demonstrate the sequencing of generated assessments using the capability taxonomy, subject taxonomy, and context derived from the competency model. This experiment considered the opinions of the students on the efficiency and effectiveness of the generated tests. The following questions were explored.

- Did the generated test fairly assess the level of students' knowledge?
- How well did the adaptive sequencing help students to identify topics that they do not know?
- How well did the adaptive sequencing help students to understand how a given learning outcome separated into learning outcome components?

- How well did the adaptive sequencing help students to understand how a given learning outcome separated into topics?
- How well did the adaptive sequencing provide a complete assessment of the level of students' knowledge?
- How useful was the adaptive sequencing for students' own self-assessment?
- Was the generated sequencing of the test semantically acceptable to an expert teacher of the domain?

The dependent variables were the student ratings on the criteria of fairly assessing the level of students' knowledge, adapting the next question, usefulness for self-assessment, identifying what students do not know, separating a given learning outcome into the learning outcome components and topics, and showing appropriately difficult questions.

The questionnaire (see appendix C) comprised twelve items as follows.

Item 1: The test fairly assesses the level of my knowledge of the following learning outcomes. This item coded as TestAssessKw.

Item 2: The system adapted the next question based on my answer. This item coded as AdaptQuestion.

Item 3: It was useful for self-assessment, when the system adapts the next question based on my answer. This item coded as UsefulforSelfAssessment.

Item 4: The system helped me to identify what I do not know based on a given learning outcome. This item coded as IdentLO.

Item 5: The system helped me to understand how a given learning outcome breaks down into learning outcome components. This item coded as DecomposeLO.

Item 6: The system helped me to understand how a given learning outcome breaks down into topics. This item coded as DecomposeTopic.

Item 7: The system showed me an appropriately difficult question based on my previous answer. This item coded as ShowDifficultQ.

Item 8: The questions in the test were clear. This item coded as QClear.

Item 9: The questions in the test fairly assessed the level of my knowledge of the following learning outcomes. This item coded as QAssessKw.

Item 10: The questions in the test were useful for further study. This item coded as QUsefulFurtherStudy.

Item 11: The questions in the test were useful for my own self-assessment. This item coded as QUsefulSelfAssessment.

Item 12: The questions in the test were appropriately difficult. This item coded as QAppropriatelyDifficult.

Students were asked to rate each item on a 4-point forced-choice Likert scale (‘Strongly disagree’, ‘Disagree’, ‘Agree’, ‘Strongly agree’, coded as 1, 2, 3, and 4 respectively) that best described their opinion.

The competencies were collected from the INFO2007 Systems Analysis and Design course. The topics of the course involved function points and associated issues as follows:

- adjusted function points
- unadjusted function points
- complexity adjustment
- the formula of the complexity adjustment factor
- degrees of influence
- the formula of the unadjusted function points
- calculating function points from an ER Diagram
- calculating function points from
    - number of attributes,
    - number of relationship lines
    - number of data entities.

Subject matter content for the competency data model was also collected from the core textbook and website of the course syllabus and reviewed by a domain expert in this field. Table 8-9 and Table 8-10 represent this data. The questions involved two distinct capabilities: Define and Calculate.

Competence	Subject Matter	Capability	Context	Sub-competence
Students can calculate adjusted function points	Concept: adjusted function points	Apply: Calculate	A customer generally asks about the availability of various car types and about the terms and costs of hire before making a reservation. When the customer collects the car, the Front Office completes the required details for the hire. Sometimes a customer will call in and collect a car without a reservation. It can also happen that, at the time of collection, the correct type of car is not available. In this case, the customer is given a superior type of car at no extra charge. When the customer returns the car, the Front Office calculates the charges passed to the Accounts Department, where invoices are raised and payments settled. The Front Office keeps records of payments it receives and passes these to the Accounts Department. Management sets the car tariffs, and requires information about the pattern of care hire at various times of year.	Students can calculate complexity adjustment
Students can calculate complexity adjustment	Concept: complexity adjustment	Apply: Calculate	(same as above)	–

Table 8-9 Some examples of function point competencies

Subject Matter	Related subject matter
Concept: adjusted function points	Concept: unadjusted function points Concept: complexity adjustment
Concept: complexity adjustment	Fact: degrees of influence Fact: the formula of the complexity adjustment factor

Table 8-10 Some examples of subject matter content based on the function points topic

Learning outcome	Generated question
Students can calculate complexity adjustment.	The formula of the complexity adjustment factor may be defined as
	Calculate complexity adjustment for this situation: A customer generally asks about the availability of various car types […] and requires information about the pattern of car hire at various times of year.
	Calculate Degrees of influence for this situation: A customer generally asks about the availability of various car types […] and requires information about the pattern of car hire at various times of year.

Table 8-11 Example generated questions

The questions were generated from the question templates as shown in Table 7-6, and examples of generated questions are shown in Table 8-11. While the system successfully generated the questions, the generated questions pointed to the critical challenges of appropriate capability and subject matter decomposition, and maintaining standards of English grammar.

The system generated 20 questions for a test on the function point topic. These questions were reduced to 14, based on two domain experts' review and their selection of the questions which would most appropriately address the experimental hypotheses. These six questions omitted, were related to some of the topics mentioned in section 8.4.2, these being 'Calculate ER Diagram', 'Calculate the formula of the complexity adjustment factor', 'Calculate the formula of the unadjusted function points', 'Define number of data entities', 'Define attributes', and 'Define relationship lines'. The rule for determining the adaptive sequencing was implemented as follows. If the student succeeds on the question, a slightly more challenging question is presented next. They would be shown the next question at the same ability level but higher subject matter level than the previous question. If the student failed the question, the system would administer successively easier questions. They would be shown the next question at the same ability level but lower subject matter level than the previous question. The choice of the first question was the highest subject matter level at top level of ability.

The participants were voluntary 2nd year undergraduate students of the INFO2007 Systems Analysis and Design course. The questionnaires were randomly distributed to all attending students at the end of a lecture. The study gathered data from 19 students.

Measured Variables	Test Value = 2.5					
	t	df	Sig. (2-tailed)	Mean Difference	95% confidence interval of the difference	
					Lower	Upper
TestAssessKw	–0.224	18	0.826	–0.026	–0.27	0.22
AdaptQuestion	5.786	18	0.000	0.711	0.45	0.97
UsefulforSelfAssessment	2.471	18	0.024	0.500	0.07	0.93
IdentLO	3.269	18	0.004	0.500	0.18	0.82
DecomposeLO	3.139	18	0.006	0.447	0.15	0.75
DecomposeTopic	0.907	18	0.376	0.184	-0.24	0.61
ShowDifficultQ	8.367	18	0.000	0.605	0.45	0.76
QClear	1.788	18	0.091	0.342	-0.06	0.74
QAssessKw	2.357	18	0.030	0.289	0.03	0.55
QUsefulFurtherStudy	3.720	18	0.002	0.447	0.19	0.70
QUsefulSelfAssessment	4.763	18	0.000	0.658	0.37	0.95
QAppropriatelyDifficult	4.595	18	0.000	0.553	0.30	0.81

Table 8-12 t Test

A one-sample t test was used to test differences between the observed sample means and the expected sample means. In this experiment, the expected mean value was assigned as 2.5, being mid-way between agreeing and disagreeing on the measurement scale. As can be seen in Table 8-13, the mean rating was significantly higher than 2.5 for 9 of the 12 measured variables. The students from the INFO2007 course agreed that:

- The system adapted the next question based on their answer ($p<0.001$).
- The adaptive assessment system was useful for self-assessment ($p=0.024$).
- The system helped them to identify what they did not know based on a given learning outcome ($p=0.004$).
- The system helped them to understand how a given learning outcome separated into learning outcome components ($p=0.006$) but they did not agree that the system helped them to understand how a given learning outcome separated into topics ($p=0.376$).
- The system showed students an appropriately difficult question based on a previous answer ($p<0.001$).
- The questions in the test fairly assessed the level of their knowledge ($p=0.030$), although, they did not agree that the whole test fairly assessed the level of their knowledge ($p=0.826$).

- The questions in the test were useful for further study ($p=0.002$).
- The questions in the test were useful for their own self-assessment ($p<0.001$) but the questions were not clear ($p=0.091$).
- The questions in the test were appropriately difficult ($p<0.001$).

Measured Variables	TestAssess Kw	Adapt Question	Usefulfor SelfAssess ment	IdentLO	Decom poseLO	Decom poseTopic	Show DifficultQ	QClear	QAssess Kw	QUseful Further Study	QUseful SelfAssess ment	QAppro priately Difficult
TestAssess Kw	1	0.224	0.000	0.162	0.257	–0.019	0.018	0.704§	0.383	0.304	0.104	0.315
Adapt Question	0.224	1	0.706§	0.311	0.035	0.617§	0.191	0.079	0.357	0.240	0.064	0.750§
UsefulforSelf Assessment	0.000	0.706§	1	0.283	0.101	0.712§	0.200	0.151	0.353	0.120	0.209	0.721§
IdentLO	0.162	0.311	0.283	1	0.537†	0.377	0.000	0.200	0.000	–0.159	0.277	0.159
Decompose LO	0.257	0.035	0.101	0.537†	1	0.372	0.030	0.305	0.466†	–0.350	0.172	0.009
Decompose Topic	–0.019	0.617§	0.712§	0.377	0.372	1	0.325	0.004	0.321	0.082	0.099	0.517†
Show DifficultQ	0.018	0.191	0.200	0.000	0.030	0.325	1	0.067	0.139	0.035	0.200	0.301
QClear	0.704§	0.079	0.151	0.200	0.305	0.004	0.067	1	0.295	0.107	0.052	0.274
QAssessKw	0.383	0.357	0.353	0.000	0.466†	0.321	0.139	0.295	1	–0.240	0.109	0.438
QUseful FurtherStudy	0.304	0.240	0.120	–0.159	–0.350	0.082	0.035	0.107	–0.240	1	–0.324	0.213
QUsefulSelf Assessment	0.104	0.064	0.209	0.277	0.172	0.099	0.200	0.052	0.109	–0.324	1	0.148
QAppropriate lyDifficult	0.315	0.750§	0.721§	0.159	0.009	0.517†	0.301	0.274	0.438	0.213	0.148	1

§ Correlation is significant at the 0.01 level (2-tailed).
† Correlation is significant at the 0.05 level (2-tailed).

Table 8-13 Pearson product-moment coefficients of correlation between student ratings

Table 8-13 provides the correlations between student ratings. The students who rated the test as fairly assessing the level of students' knowledge also rated the questions as being clear.

The students who rated the system as being adaptive also rated the system as being useful for self-assessment; as helping them to understand how a given learning outcome separated into topics; and they also rated the questions in the test as being appropriately difficult.

The students who rated the system as being useful for self-assessment also rated the system as being adaptive; as helping them to understand how a given learning outcome separated into topics; and they also rated the questions in the test as being appropriately difficult.

The students who rated the system as helping them to identify what they did not know also rated the system as helping them to understand how a given learning outcome separated into learning outcome components.

The students who rated the system as helping them to understand how a given learning outcome separated into learning outcome components also rated the system as helping them to identify what they did not know and the questions in the test as fairly assessing the level of their knowledge.

The students who rated the system as helping them to understand how a given learning outcome separated into topic also rated the system as being adaptive and useful for self-assessment, and the questions in the test as being appropriately difficult.

The students who rated the question as being clear also rated the test as fairly assessing the level of their knowledge.

The students who rated the question as fairly assessing the level of their knowledge also rated the system as helping them to understand how a given learning outcome separated into learning outcome components.

The students who rated the questions as being appropriately difficult also rated the system as being adaptive and useful for self-assessment and as helping them to understand how a given learning outcome separated into topics.

All other correlations were insignificant.

Figure 8-5 provides the cluster analysis which groups the questionnaire answers using Ward's method. Ward's hierarchical clustering method tends to create clusters of small size to minimise the loss associated with each grouping (Field, 2005).

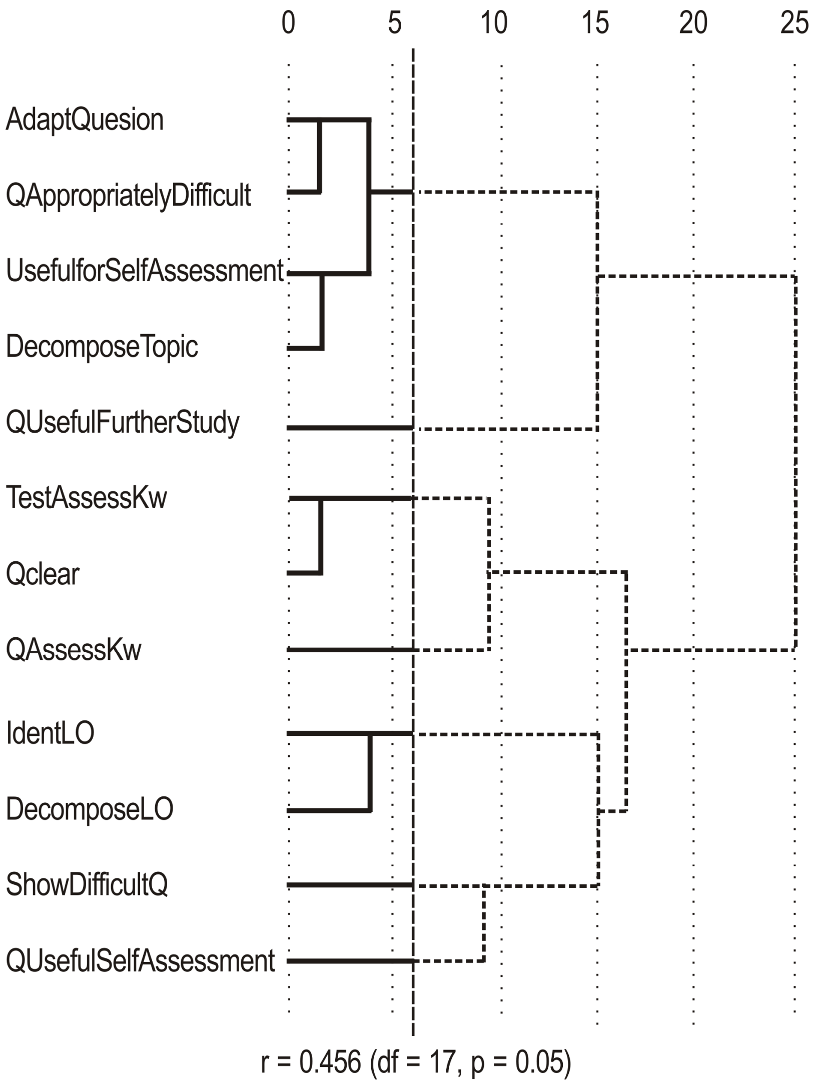


Figure 8-5 Dendrogram of Hierarchical Cluster analysis using Ward method

Figure 8-5 shows that "AdaptQuestion" & "QAppropriatelyDifficult", "UsefulforSelfAssessment" & "DecomposeTopic", "TestAssessKw" & "QClear", and "IdentLO" & "DecomposeLO" were the three clusters which formed first.

A cut-off line for approximating cluster significance may be set at the critical value of the Pearson correlation coefficient r = 0.456 (df =17, p=0.05). Using this line, Figure 8-5 illustrates the resulting clusters as follows.

Cluster 1 comprises 4 questions which are "AdaptQuestion", QAppropriatelyDifficult", "UsefulforSelfAssessment", and "DecomposeTopic".

Cluster 2 comprises 2 questions which are "TestAssessKw" and "QClear".

Cluster 3 comprises 2 questions which are "IdentLO" and "DecomposeLO".

Four questions (QUsefulFurtherStudy, QAssessKw, ShowDifficultQ, QUsefulSelfAssessment) did not join any cluster.

8.4.3 Experiment 3

The third experiment was developed to compare the pedagogical effectiveness of a variety of adaptive sequences by analysing the opinions of students and experts on two adaptive sequencing algorithms and two starting points for questioning. The following questions were explored.

- Which sequences do students think more pedagogically effective?
- Which sequences do experts think more pedagogically effective?

This experiment was based on the second experiment. The competencies were those incorporated in the second experiment. The system generated questions for four different types of test corresponding to the two generating algorithms and the two starting points for questioning.

The independent variables were the algorithm and the starting point. For the first algorithm, if the student succeeded on a question, a slightly more challenging question was presented next. This was a question at the same capability level and at a higher subject matter level than the previous question. If the student failed the question, the system presented an easier question. This was a question at the same capability level and at the same or a lower subject matter level than the previous question. This algorithm was called "SameAbilityLevel".

For the second algorithm, if the student succeeded on a question, a slightly more challenging question was presented next. This was a question at the same or a higher capability level and at the same or a higher subject matter level than the previous question. If the student failed the question, the system presented an easier question as in the first algorithm. This was a question at the same or a lower capability level but at a lower subject matter level than the previous question. This algorithm was called "HigherAbilityLevel"

One starting point presented the first question from the lowest subject matter level and the lowest ability level. This starting point was called "Bottom-up". Another starting point presented the first question from the highest subject matter level and the highest ability level. This starting point was called "Top-down".

Experiment 3.1

This experiment was developed to explore sequences which students think are more pedagogically effective between four different types of test. The students were asked to compare the sequences according to the different starting points with the same algorithm.

The dependent variables were the student ratings on the six criteria and their percentage correct. The questionnaire (see appendix D) comprised seven items as follows.

Item 1: What was your percentage of correct answers? This item coded as PercentageCorrect.

Item 2: The test fairly assesses the level of my knowledge of the following learning outcomes. This item coded as TestAssessKw.

Item 3: How well did the adaptive sequencing help you to identify what you do not know? This item coded as IdentKw.

Item 4: How well did the adaptive sequencing help you to understand how a given learning outcome separated into learning outcome components? This item coded as HelpLO.

Item 5: How well did the adaptive sequencing help you to understand how a given learning outcome separated into topics? This item coded as HelpTopic.

Item 6: How well did the adaptive sequencing provide a complete assessment of the level of your knowledge? This item coded as CompleteAss.

Item 7: How useful was the adaptive sequencing for your own self-assessment? This item coded as UsefulSelf.

The questionnaire asked the students to rate each item on a 5-point Likert scale (‘Totally poor’, ‘Poor’, ‘Acceptable’, ‘Pretty good’, ‘Wonderful’, coded as 1, 2, 3, 4, and 5 respectively) that best described their opinion of the six opinion items. In addition, the students noted the percentage of their answers that were correct. The study population was divided into two groups, where each group was presented with a different adaptive sequencing algorithm. For each group, half were assigned the ‘Top-down’ starting point and half the ‘Bottom-up’ starting point. For their second test, they commenced from the other starting point. This ordering controlled for practice effects.

The participants were voluntary 3rd year undergraduate students of the INFO3004 eLearning and Learning Technology course. The questionnaires were randomly distributed to all attending students at the end of a lecture. Each questionnaire comprised 14 items (14 measured variables). The first 7 variables related to the first test administered and the second 7 variables related to the second test administered. The study gathered data from 21 students.

Univariate tests were used to analyse the data from this experiment in order to evaluate the interaction effects between the independent variables. Usually, confidence intervals are calculated at 95% confidence and test statistics evaluated at an alpha level of 0.05, but for this exploratory experiment, 90% confidence intervals were used and an alpha level of 0.10 was adopted.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Order	423.921	1	423.921	0.612	0.440
Algorithm	973.090	1	973.090	1.404	0.244
StartingPoint	82.160	1	82.160	0.119	0.733
Order * Algorithm	1122.088	1	1122.088	1.619	0.212
Order * StartingPoint	5282.139	1	5282.139	7.620	0.009
Algorithm * StartingPoint	1467.257	1	1467.257	2.117	0.155
Order * Algorithm * StartingPoint	1009.738	1	1009.738	1.457	0.236
Error	23568.789	34	693.200		

Table 8-14 Tests of Between-Subjects Effects for PercentageCorrect

As can be seen in Table 8-14, the 3rd order interaction effects (Order * Algorithm * StartingPoint), 2nd order interaction effects (Order * StartingPoint, Order * Algorithm), and 1st order effects (Algorithm) were insignificant on percentage correct. There was a significant 2nd order interaction effect (Order * StartingPoint) on percentage correct ($F=7.620$, df =1, 34, $p<0.10$).

Order	Starting Point	Mean	Std. Error	90% Confidence Interval	
				Lower Bound	Upper Bound
Top Down First	Top Down	33.861	7.971	20.382	47.340
	Bottom Up	59.166	7.971	45.687	72.645
Top Down Second	Top Down	49.986	8.326	35.908	64.064
	Bottom Up	30.293	8.326	16.215	44.371

Table 8-15 Estimated Marginal Means of PercentageCorrect for starting point

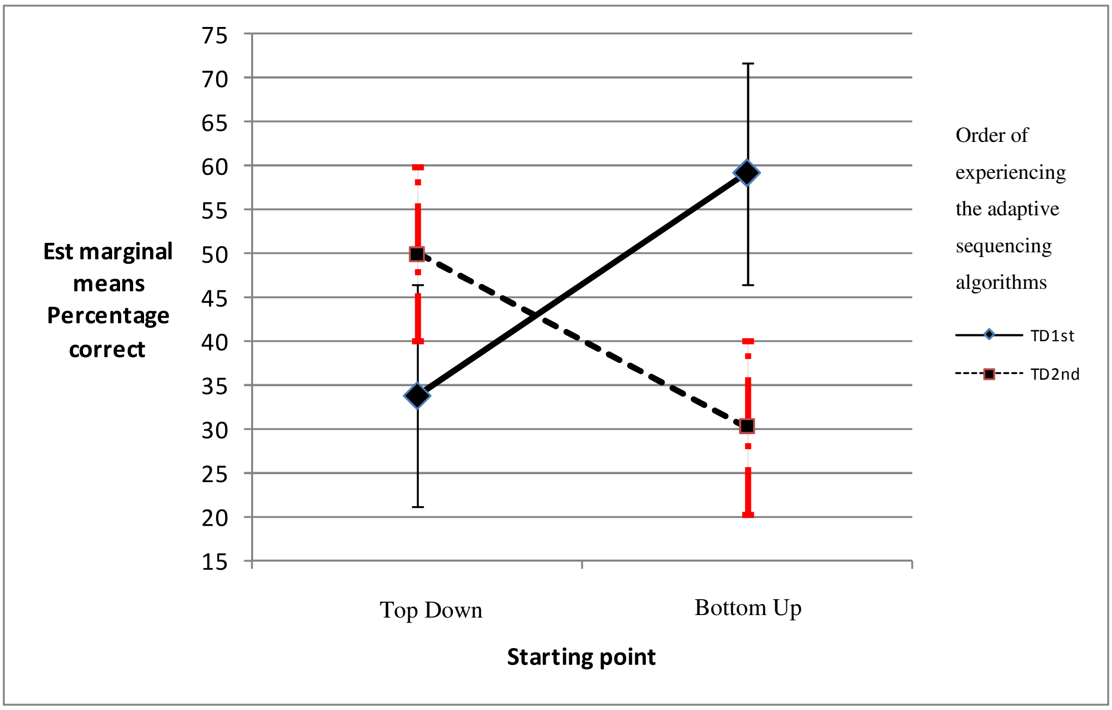


Figure 8-6 Means profile plots of estimated marginal mean of PercentageCorrect for starting point

Table 8-15 provides the estimated marginal means according to starting point, to test whether there is significantly different mean percentage correct between starting points, and Figure 8-6 shows the profile graphs. An inspection of the profile graphs and reference to the confidence intervals of Table 8-15 shows that

- Mean percentage correct for the students experiencing the Top-down method was significantly lower than for those experiencing the Bottom-up method, when they started with the Top-down method.
- Mean percentage correct for the students experiencing the Top-down method was significantly higher than for those experiencing the Bottom-up method, when they started with the Bottom-up method.
- Mean percentage correct for the students starting with the Top-down method was not significantly different from starting with the Bottom-up method.
- Mean percentage correct for the students starting with the Bottom-up method was significantly lower than starting with the Top-down method.

Appendix E shows that the 3rd order interaction effects (Order * Algorithm * StartingPoint), 2nd order interaction effects (Algorithm * StartingPoint, Order * StartingPoint, Order * Algorithm), and 1st order effect (Order, Algorithm, StartingPoint) were all not significant, indicating significant differences between the algorithm, starting point and order, of students' opinions on what they did know, understanding how a given learning outcome decomposed into learning outcome components, understanding how a given learning outcome decomposed into topic, and providing a complete assessment of the level of their knowledge.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Order	4.522	1	4.522	3.693	0.063
Algorithm	.018	1	0.018	0.015	0.904
StartingPoint	.696	1	0.696	0.569	0.456
Order * Algorithm	.696	1	0.696	0.569	0.456
Order * StartingPoint	1.340	1	1.340	1.094	0.303
Algorithm * StartingPoint	0.696	1	0.696	0.569	0.456
Order * Algorithm * StartingPoint	0.018	1	0.018	0.015	0.904
Error	41.633	34	1.225		

Table 8-16 Tests of Between-Subjects Effects for UsefulSelf-Assessment

As can be seen in Table 8-16 , the 3rd order interaction effects (Order * Algorithm * StartingPoint), 2nd order interaction effects (Order * StartingPoint, Order * Algorithm, Algorithm * StartingPoint), and 1st order effect (Algorithm, StartingPoint) were not significant on usefulness of the adaptive sequencing for self-assessment. There was a significant 1st order effect (Order) on usefulness of the adaptive sequencing for self-assessment (F=3.693, df =1, 34, p<0.10).

Order	Mean	Std. Error	90% Confidence Interval	
			Lower Bound	Upper Bound
TopDownFirst	3.308	0.237	2.827	3.790
TopDownSecond	2.650	0.247	2.147	3.153

Table 8-17 Estimated Marginal Means for UsefulSelf-Assessment

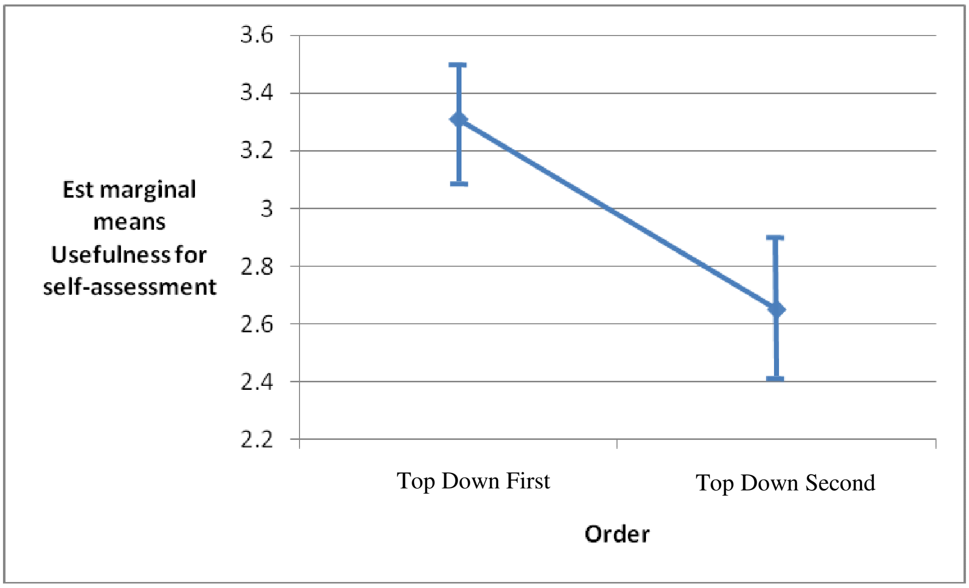


Figure 8-7 Means profile plots of estimated marginal mean for order

Table 8-17 provides the estimated marginal means for the order, to examine whether there is significantly different mean usefulness for self-assessment according to the ordering, and Figure 8-7 shows the profile graphs. An inspection of the profile graphs and reference to the confidence intervals of Table 8-17 shows that mean usefulness of self-assessment for the students experiencing the Top-down method first was significantly higher than for those experiencing the Bottom-up method first.

Experiment 3.2

This experiment was developed to explore question sequences which experts think are more pedagogically effective. The participants in this experiment were experts who have some experience with the 'Function points' topic at School of Electronics and Computer Science, University of Southampton.

Four tests were generated using the two algorithms and two starting points as explained in Experiment 3. The resulting sequences of questions were labelled test A (algorithm: SameAbilityLevel, starting point: Top-down), test B (algorithm: SameAbilityLevel, starting point: Bottom-up), test C (algorithm: HigherAbilityLevel, starting point: Top-down), and test D (algorithm: HigherAbilityLevel, starting point: Bottom-up).

A pilot study was designed to test logistics and gather information prior to Experiment 3.2 and to refine the design of the experiment. A questionnaire asked a pilot participant to compare all four sequences together at one time. The expert thought that the comparisons of all four at once resulted in an excessive cognitive load and made comparative decisions difficult and very slow. As a result, the questionnaire was redesigned (see appendix F). Sequences were compared in pairs rather than comparing all four sequences together.

In this experiment, the experts were asked to compare the sequences in pairs. Six sequence pairs were constructed by combining each test with every other test. The order of presentation of the sequence pairs to the experts was randomised to control for practice effects. For example, the first expert would be assessed when comparing tests A&B, A&C, A&D, B&D, C&D, and B&C. The second expert would compare tests A&C, A&D, B&D, C&D, B&C, and A&B.

The experts compared a pair of tests using a questionnaire, which comprised six items as follows.

Item 1: Which test more fairly assesses the level of a student's knowledge? This item coded as TestAssessKw.

Item 2: Which test more helps a student to identify what they do not know? This item coded as IdentKw.

Item 3: Which test helps a student more to understand how a given learning outcome separated into learning outcome components? This item coded as HelpLO.

Item 4: Which test helps a student more to understand how a given learning outcome separated into topics? This item coded as HelpTopic.

Item 5: Which test provides a more complete assessment of the level of a student's knowledge for a teacher? This item coded as CompleteAss.

Item 6: Which test more usefully provides for a student's own self-assessment? This item coded as UsefulSelf.

A measure was calculated for a test by scoring it '1' if it was rated better than its compared test, or '0' if not. These scores were added up to yield a measure of the pedagogical effectiveness of the test when compared with every other test. The measurement range was thus 0-3. For example, A:2, B:1, C:0, and D:3 indicated that 2 out of 3 pairings for test A rated it more pedagogically effective. Note that the total of the measures for the four tests was always 6.

In this experiment, the independent variables were the algorithm and the starting point. The dependent variables were the expert measures on the six criteria. The study gathered data from 5 experts.

Effect	The statistic method	Value	F	Hypoth df	Error df	Sig.
Algorithm	Wilks's Lambda	0.373	3.085	6	11	0.050
	Hotelling's Trace	1.683	3.085	6	11	0.050
StartingPoint	Wilks's Lambda	0.631	1.074	6	11	0.433
	Hotelling's Trace	0.586	1.074	6	11	0.433
Algorithm * StartingPoint	Wilks's Lambda	0.724	0.700	6	11	0.656
	Hotelling's Trace	0.382	0.700	6	11	0.656

Table 8-18 Multivariate test between the four sequences of questions with algorithm and starting point

As can be seen in Table 8-18, the multivariate tests for differences in measures according to algorithm and starting point, and the algorithm by starting point interaction, showed significance only for differences between algorithm (Wilks's Lambda p = 0.05 and Hotelling's Trace p = 0.05).

Dependent Variable	Algorithm	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
TestAssessKw	SameAbilityLevel	1.100	0.361	0.336	1.864
	LowerAbilityLevel	1.900	0.361	1.136	2.664
IdentKw	SameAbilityLevel	1.000	0.335	0.289	1.711
	LowerAbilityLevel	2.000	0.335	1.289	2.711
HelpLO	SameAbilityLevel	0.800	0.304	0.155	1.445
	LowerAbilityLevel	2.200	0.304	1.555	2.845
HelpTopic	SameAbilityLevel	0.900	0.332	0.197	1.603
	LowerAbilityLevel	2.100	0.332	1.397	2.803
CompleteAss	SameAbilityLevel	1.100	0.312	0.438	1.762
	LowerAbilityLevel	2.100	0.312	1.438	2.762
UsefulSelf	SameAbilityLevel	1.000	0.335	0.289	1.711
	LowerAbilityLevel	2.000	0.335	1.289	2.711

Table 8-19 Estimated Marginal Means for algorithm

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Algorithm	TestAssessKw	3.200	1	3.200	2.462	0.136
	IdentKw	5.000	1	5.000	4.444	0.051
	HelpLO	9.800	1	9.800	10.595	0.005
	HelpTopic	7.200	1	7.200	6.545	0.021
	CompleteAss	5.000	1	5.000	5.128	0.038
	UsefulSelf	5.000	1	5.000	4.444	0.051
Error	TestAssessKw	20.800	16	1.300		
	IdentKw	18.000	16	1.125		
	HelpLO	14.800	16	0.925		
	HelpTopic	17.600	16	1.100		
	CompleteAss	15.600	16	0.975		
	UsefulSelf	18.000	16	1.125		

Table 8-20 Tests of Between-Subject Effects between algorithm and six variables

Table 8-19 provides the estimated marginal means for the six measures according to algorithm. Table 8-20 provides the test of between-subject effects for algorithm, where it may be seen that there were significant differences in mean measures of 'IdentKw', 'HelpLO', 'HelpTopic', 'CompleteAss', and 'UsefulSelf', but there was no significant differences in mean measures of 'TestAssessKw'.

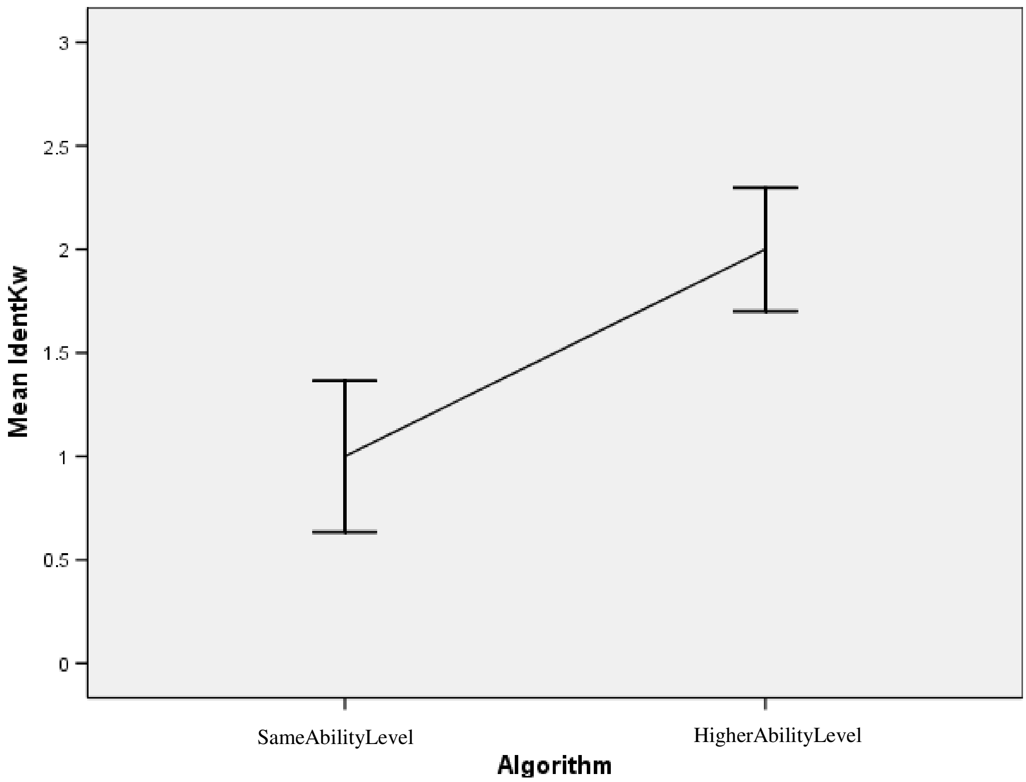


Figure 8-8 Means profile plots of estimated marginal mean IdentKw for algorithm

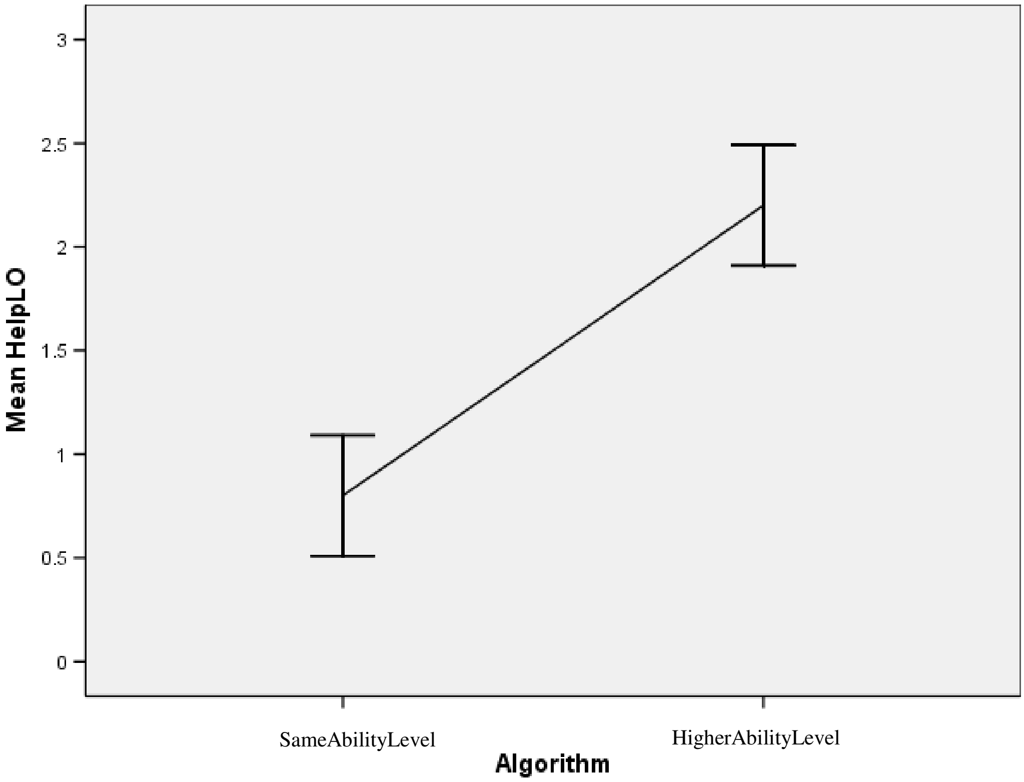


Figure 8-9 Means profile plots of estimated marginal mean HelpLO for algorithm

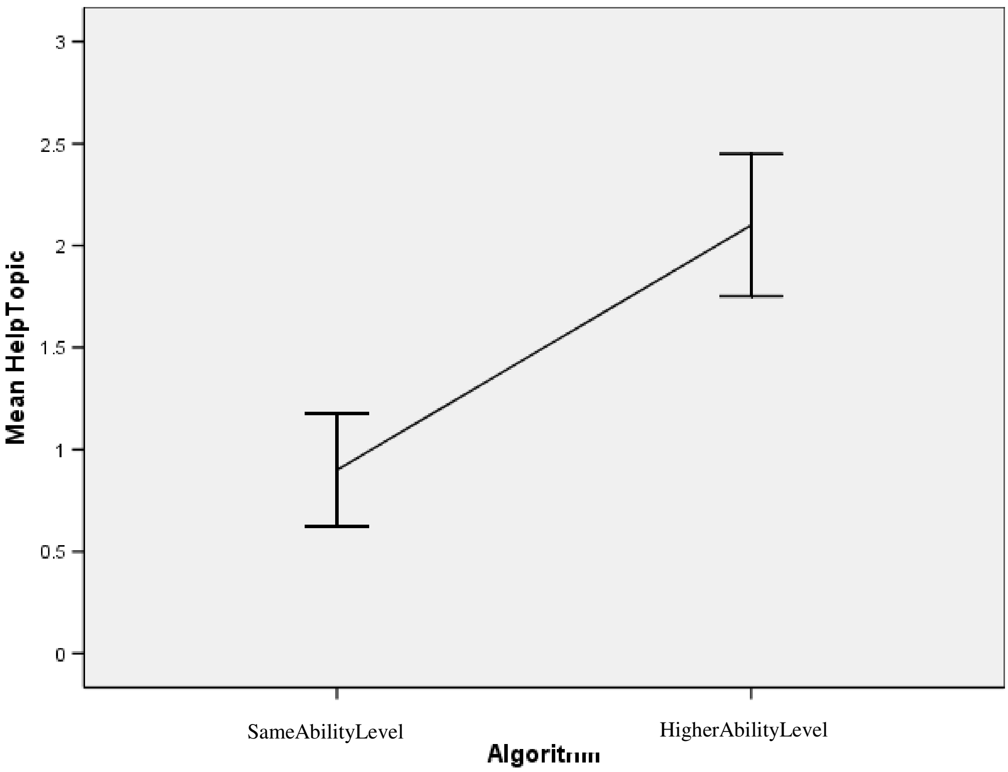


Figure 8-10 Mean profile plots of estimated marginal mean HelpTopic for algorithm

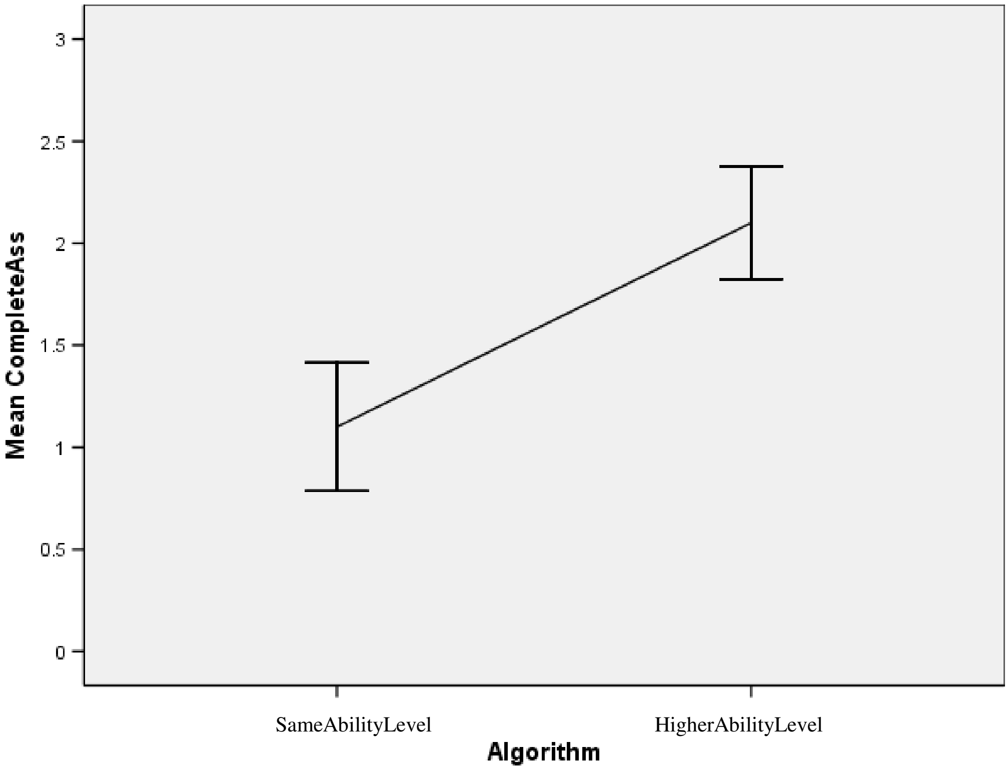


Figure 8-11 Mean profile plots of estimated marginal mean CompleteAss for algorithm

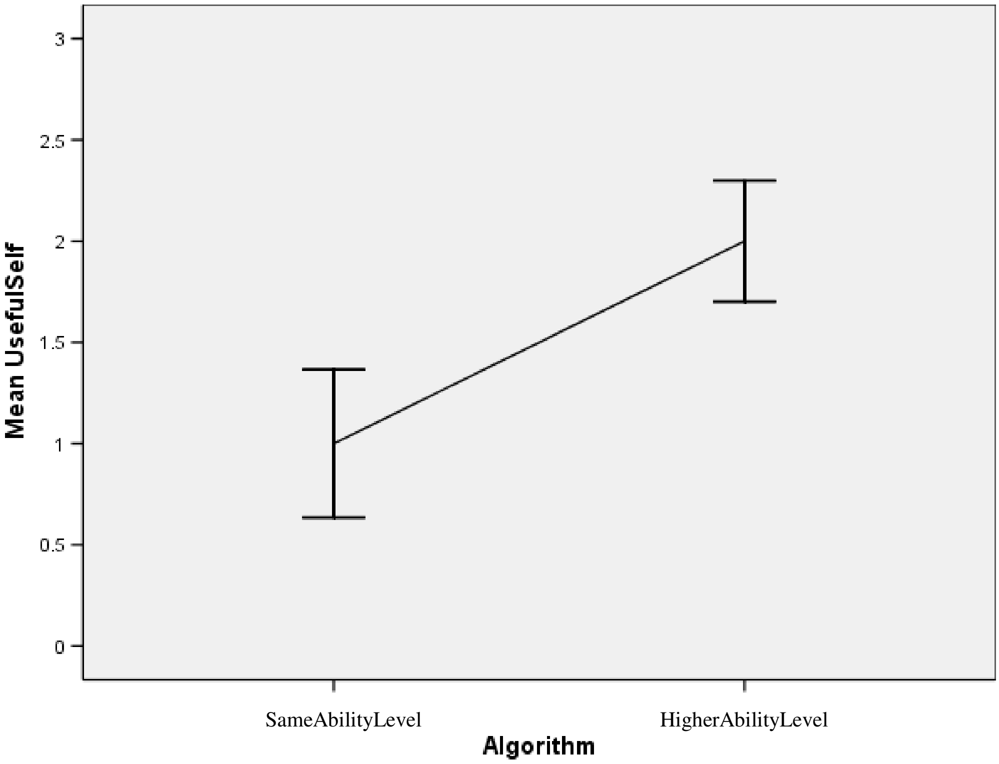


Figure 8-12 Mean profile plots of marginal mean UsefulSelf for algorithm

The profile graphs in Figure 8-8, Figure 8-9, Figure 8-10, Figure 8-11, and Figure 8-12, show the estimated marginal mean with upper and lower intervals of one standard error. Reference to these figures and to the confidence intervals of Table 8-19 show that the mean IdentKw, HelpLO, HelpTopic, CompelteAss, and UsefulSelf for 'HigherAbilityLevel' algorithm was significantly higher than for the 'SameAbilityLevel' algorithm. There was no significant difference in mean measures for TestAssessKw.

8.5 Summary

This chapter presented the results of experiments that report the efficiency and effectiveness of the generated questions and generated test, and the results of comparing the effectiveness of two adaptive sequencing algorithms and two starting points of the first question on their pedagogical effectiveness. The following chapter discusses the results of the experiments.

Chapter 9 Discussion and Evaluation

9.1 Introduction

The output of the research implementation was evaluated by considering the quality of the generated questions, the usefulness and adaptiveness of the COMBA generated test, the opinion of students and experts about the adaptive sequencing of the COMBA generated test, and the relationship between a variety of adaptive sequences and their pedagogical effectiveness. Experimental results are discussed.

9.2 Experimental results

9.2.1 Experiment 1

The results indicate that the generated questions were of acceptable value to the students. The student ratings showed the specific questions were more useful, and the generic questions were more challenging. This finding suggests that the students did not enjoy answering with definitions and explanations, and preferred questions with a variety of specific situations.

The finding that both types of question did not differ significantly on the two other criteria, their clarity and whether they matched the intended learning outcomes, is not unexpected. Interestingly, there was no effect of capability type, and no interaction between capability type (define, explain, and calculate) and question type (specific and generic), indicating that ratings were similar for the three capability types. The diversity of capability type was limited. This point suggests the need to explore creative use of question styles and capability vocabularies in order to examine interaction effects between capability type and question type. Questions such as "what", "who", "when", "where", "why", may provide for new question styles which include more challenging capability vocabulary such as 'analyse' and 'synthesis'.

The authoring question template, used as the starting point in formulating the format of questions, exhibited a rather low efficiency of 59.5% (the number of the generated questions, 42, in relation to the number of selected questions, 25). The method of generating a question used parameterised templates and retrieved subject matter and capability nodes relevant to the submitted competency. Only four unsophisticated templates were implemented. Each comprised just two or three attributes which could be retrieved from the nodes. The ontology representing these nodes was based on OWL Lite, which was sufficiently expressive to describe the subject matter hierarchy, but this approach did not allow retrieving any topic which was not directly represented in the intended learning outcome and the subject matter involved.

It may be possible to use some natural language processing for developing the format of questions, and this point will be discussed in future work (Chapter 10). In addition, the COMBA domain ontology may be represented by using OWL DL, or OWL Full, to improve the description of the subject matter hierarchy.

OWL DL and OWL Full allow arbitrary Boolean combinations of classes and restrictions: unionOf, complementOf, and intersectionOf. For example, in this study, using unionOf, a "Concept" class may contain things that are either "Fact" or "Concept". From this unionOf statement, a reasoner can derive any topic which was not directly relevant in the intended learning outcome and the subject matter involved.

9.2.2 Experiment 2

As was found in Experiment 1, the results indicate that the generated test and questions were of acceptable value to the students. Five out of seven measured variables for the test, and four out of five measured variables for the questions, were rated as "Agree".

Interestingly, from the t-test, the students agreed that the system helped them to understand how a given learning outcome separated into "learning outcome components", but they did not agree that it helped them to separate a given learning outcome into "topics". Whilst a learning outcome component involves capability and subject matter, a topic involves only subject matter. This suggests that the generated questions helped the students to understand the decomposition of capability, but were not particularly helpful in understanding the decomposition of topics.

From the dendogram, the perceived adaptivity of the next question was associated primarily with presenting appropriately difficult questions. The perceived usefulness of a test for self-assessment was associated primarily with help in understanding how a given learning outcome separated into "topics". The perceived usefulness of a test for self-assessment reflected the perceived adaptivity of the next question.

In addition, the perceived fairness of the test related only to the clarity of the questions and not to any other variables, giving an interesting insight into the students' views of the importance of clarity for fairness.

Finally, the students who rated the system as helping them to identify what they did not know, also rated the system as helping them to understand how a given learning outcome separated into "learning outcome components", but they did not rate the system as helping them to understand how a given learning outcome separated into "topic". This supports the finding mentioned earlier that the system's help with the decomposition of "capability" was of greater value to the students than decomposition of "topic".

The efficiency of using the authoring question template in this experiment was improved from 59.5% to 70% (the number of the generated questions, 20, in relation to the number of selected questions, 17). The revised querying algorithms reduced some inappropriate format of questions by using revised SPARQL queries to expand the returned results. In Experiment 1, the following query was implemented.

```
SELECT ?SMC ?relatedSMC ?capability ?imply ?context
WHERE {?competence Comp:SMC ?SMC
            ?competence Comp:capability ?capability
  ?competence Comp:Context ?context
?SMC ?relationship ?relatedSMC
{?capability Comp:imply ?imply} UNION
{?competence Rdf:type Comp:Competence}};
```

Table 9-1 An example of the SPARQL query in Experiment 1

The query shown in Table 9-1 asks for "all subject matter, relevant subject matter, capability, lower capability level, and context", against desired properties in the

'WHERE' clause. In Experiment 2, this query was divided into two queries as shown in Table 9-2 and Table 9-3 to reduce redundancy of each row in the result set.

```
SELECT ?SMC ?relatedSMC ?imply
WHERE {?competence Comp:SMC ?SMC
            ?competence Comp:capability ?capability
?competence Rdf:type Comp:Competence
?capability Comp:imply ?imply
?SMC ?relationship ?relatedSMC};
```

Table 9-2 An example of the first revised SPARQL query in Experiment 2

```
SELECT ?SMC ?relatedSMC ?capability ?context
WHERE {?competence Comp:SMC ?SMC
            ?competence Comp:capability ?capability
  ?competence Comp:Context ?context
?competence Rdf:type Comp:Competence
?SMC ?relationship ?relatedSMC};
```

Table 9-3 An example of the second revised SPARQL query in Experiment 2

The query shown in Table 9-2 asks only for "all subject matter, relevant subject matter, and lower capability level", against desired properties. In Table 9-3, the query also asks for "all subject matter, relevance subject matter, capability, and context", against desired properties.

The revised querying eliminates some inappropriate format of questions which do not make a good sense or show poor grammar and syntax. This indicates that not only the format of the template itself is important for generating questions using parameterised templates, but also the algorithm of querying is critical.

9.2.3 Experiment 3

The mean percentage correct for the students of the second test was higher than that of the first test. The cause is that the students had experience with the test. In addition, the results indicate that the order of experiencing the adaptive sequencing algorithm (the students who started first with Top-down method; the students who started first with Bottom-up method) and the starting point for questioning (Top-down method, Bottom-up method) were significant effects for the percentage of correct answers.

As mentioned in Chapter 8, 'Top-down' starting point presented the first question from the highest subject matter level and the highest ability level. 'Bottom-up' starting point presented the first question from the lowest subject matter level and the lowest ability level. The results show that the average percentage of correct answers for the students experiencing the 'Top-down' method was higher than for the 'Bottom-up' method. Presenting the first question from the highest subject matter level and the highest ability level increases the percentage of correct answers for the students. This suggests that starting with a broad range, and finishing with a narrow range, of content-related questions may be more pedagogically effective.

The student ratings on their perception of the fairness of tests, identification of what they do not know, understanding how a given learning outcome separated into learning outcome components and topics, providing a complete assessment, and usefulness for self-assessment, showed no significant differences across the four sequences. This suggests that comparing the pedagogical effectiveness of adaptive sequences should be measured from the student's performance instead of the student's perception.

Interestingly, the effect of starting point and algorithm on the student ratings were not significant, while the effect of algorithm was significant for the expert ratings. This shows that in the experts' opinion, the pedagogical effectiveness of sequences depends on algorithm. The starting point had an actual effect on the student's performance but their ratings did not show difference due to the starting point. This shows that the actual effect of starting point and algorithm on the student's performance is in contrast to their perception of such starting point and algorithm.

The results suggest that the ways students perceive their learning and the ways experts perceive their teaching are dependent on their perspective, objective of assessing, and experiences.

For experts' rating, five out of six measured variables for 'HigherAbilityLevel' algorithm were significantly higher than for 'SameAbilityLevel' algorithm. This implies that the 'HigherAbilityLevel' algorithm was of more value to the experts for its pedagogical effectiveness.

According to the results of the two experiments with students and experts, the 'Top-down' method for starting point and the 'LowerAbilityLevel' algorithm would be an optimal combination for adaptive assessment.

9.3 Summary

The experiments have confirmed the potential of machine-processable competency models. The results support the capability of the COMBA system for accommodating competencies, linking competencies, and tracking the knowledge state of students. The results demonstrate the feasibility of automatically generating questions from a competency framework. The system also demonstrates new ways of assessing by measuring the knowledge and skills based on the intended learning outcomes involved. The following Chapter describes the current contribution and suggests future work that can improve the formats of questions, and potential extensions for research areas of the Semantic Web, e-learning, and Competency.

Chapter 10 Contributions and Future work

10.1 Introduction

In this chapter, contributions of a competency model, of a competency tree, and of applying taxonomy and ontology with e-assessment for generating questions and sequences, are considered. Future work arising from the design and development of the system includes enhancing the efficiency of the authoring question templates, extracting ontological vocabulary with Natural Language Processing, and integrating competence development with human resource management.

10.2 Contributions

10.2.1 A machine-processable competency model

A competency model (COMpetence-Based learner knowledge for personalized Assessment, COMBA) was developed because of the unsatisfactory results delivered by existing competency standards and taxonomies of competence. COMBA combines the concepts of subject matter and cognitive ability with other considerations such as contexts, situations, and tools.

COMBA is a structural model for machine processing using the Web Ontology Language (OWL) to express competencies. This approach addresses many of the problems of extending and combining structured content in different formats from different schemas. The use of the competency model and ontologies overcomes limitations in interoperability, portability, and reusability. The COMBA model supports consistency checking, assessing differences in knowledge levels, and comparing achievement in related domains, which were essentially impractical previously.

The key contributions of this model are that it identifies: ways to define competencies of individual learners, prerequisites and goals for resource content, the learner's knowledge state, and personalization capability. This model has the great advantage of providing individuals with a more detailed identification of learners' performance. The model could be used in conjunction with a development discussion between the learner and teacher to provide focus on the key aspects to be developed for each competency. It is suggested that information about competencies should form the basis of pedagogically-informed metadata which would be relevant to any description of content or process in a learning and teaching situation.

The pedagogical advantages of the model include generating valid and reliable questions, providing a framework for organising, planning, and designing sequences of question and learning materials, and selecting distractors based on pedagogical methods.

10.2.2 A competency tree

An assessment for a competency often actually tests component competencies. The COMBA model supports assembling these competencies and their linkages into trees. A tree structure is a particular way of representing a structure in a graphical form (Johnson and Shneiderman, 1991). While the relationship between nodes is modelled as a family relation such as parent and child, there is no ordering of nodes on the same level, and this yields a tree structure rather than a hierarchy. It is assumed that all children of a defined competency are required in order to achieve proficiency at the parent level. While the tree structure defines a top-down or bottom-up structure, it does not imply sequencing as might be implied in a hierarchy. For example, a competency tree may specify how to roll up the assessment for each competency throughout a competency tree without implying sequencing of assessments of same level competencies. One of the advantages of a competency tree structure is that it separates the composition rule in the domain from other structural components.

The key contribution is that a competency tree helps guide and inform practice for evaluation and navigation. This would help teachers think about and identify ways of improving and evaluating assessment systems. A competency tree can provide a framework for organising, planning, and designing sequences of questions and learning materials to guide learning and teaching.

Using the competency tree in adaptive assessment requires rules, perhaps based on pedagogical principles, in order to sequence and control the adaptive process. The tree structure can help visualise complicated information and reveal relationships that might otherwise be hidden. This approach allows teachers to develop adaptive instruction and assessments that is effective and appealing for many groups of learners.

10.2.3 Taxonomies as a structural basis for e-assessment

Questioning reveals learners' abilities to reason, create, analyse, synthesise, and evaluate. Effective questions can promote the thinking and performance of learners. However, creating effective questions is time-consuming because it may require considerable resources and skills in critical thinking. The questions have to be carefully defined in order to accurately represent the intended learning outcome and the subject matter content involved.

In this study, subject matter and cognitive ability taxonomies in the competency model were used to create a list of all questions that are possible at various levels. These questions were used to test understanding and in some cases determine the degree to which learners had actually acquired the desired knowledge. The competency model guides question development through the six levels of cognitive ability to stimulate critical thinking. This makes it possible to guide learners in developing questions for themselves and provide authoring templates to speed the creation of new questions for self-assessment.

10.2.4 Ontologies as a semantic basis for e-assessment

Currently, e-assessment systems lack integration, interoperability, portability, and reusability with other systems and environments. Ontologies support automation, integration, and reuse of data across diverse applications. An ontology is an explicit and formal specification for the description of the main concepts of a domain and their relationships, thus providing a machine-processable shared understanding of a domain.

For the COMBA model, the ontology supports connecting resources available in a domain and representing knowledge states of learners. Ontological metadata expresses terms defined formally and unambiguously. This metadata provides information for e-assessment in order to integrate and reuse these data with other systems, and for adaptive assessment systems in order to adapt their behaviour and structure according to the personal needs and ability of each learner. Structuring knowledge in a new domain by using the conceptualization in ontologies should allow faster build of new systems.

Using ontologies allows for creative uses of content in novel ways. For example, the use of ontologies can enable an implementation of intelligent software agents helping the learner to find and use globally distributed learning resources; collaborative and distributed authoring and course construction; and reuse of learning material for future study. For browsing or searching, the metadata within an ontology can assist an intelligent search engine to process a query by automatically generalising the query to find nearest partial matches instead of returning no results.

10.2.5 Semi-automatic generation of questions

This study presented one possibility for using ontologies in automating question generation. The model enables an assessment item to be formulated directly from a competence specification by using question templates and the parameters of that competence: capability, subject matter content, and other elements such as the situation. The templates are designed to have the structure of a well constructed question with the parameters. Each parameter enables the generation of a series of questions within the same template.

The key contribution of this approach is that it demonstrates ways to generate questions semi-automatically from a competency tree. It is suggested that automatic generation of questions using parameterised templates could exploit a competency ontology model which provides an alternative to the lengthy and demanding activity of developing effective questions. In assisting developers to produce questions in a fast and expedient manner without compromising quality, the use of automatic generation of questions saves both time and production costs. This methodology is general and can be extended to other fields too.

10.2.6 Various methods for generating adaptive question sequences

The competency tree could be used to drill down into the component competencies of a target competency, helping to define what to assess and how to assess it. A competency tree could support a variety of adaptive rules to adjust questions to the student's capability and to the nature of their knowledge. Many methods of traversing the competency tree could be applied, involving different starting points and algorithms. These methods may lead to interesting issues which should be considered in adapting to the learners' particular talents, strengths, weakness, and own learning preferences.

The key contribution is supporting a variety of ways of developing adaptive sequences. This study demonstrated four methods for generating adaptive question sequences and considered their pedagogical value. For example, it is possible that students might have differing abilities in quite similar content areas. In this case, learners may not achieve an appropriate level of their capability and content. New adaptive question sequences could employ different traversal algorithms. If the learner failed a question, the system could present the next question at a lower capability level and at the same subject matter level; or at the same capability level and at the nearest subject matter level to the previous question. The pedagogical value of a particular method would need further investigation for successful learning and teaching, but having such varieties of methods could provide fruitful areas of exploration.

10.2.7 Generating distractors

One of the main challenges in generating multiple choice questions is the provision of plausible distractors. A competency tree allows the selection of plausible distractors derived from nodes semantically close to the 'correct' node from the tree. This would make each distractor similar to the correct answer, as well as consistent with the key concepts of the question. The methodology of selecting distractors can be based on pedagogical methods by adapting the traversal algorithm. For example, distractors can be selected from unfamiliar words in context, requiring learners to make inferences.

Each generated distractor is constructed from nodes of the tree which can represent plausible and common errors that a learner might make. When generated from the competency tree that reflects levels of content taxonomy and capability taxonomy, these distractors enable the development of a rich breadth and depth of multiple choice questions.

Using such questions, teachers can contribute to an analysis of a learner's pattern of misunderstanding in the subject area. The competency tree allows the question to have distractors spread across all level of a content taxonomy, thereby helping the teacher identify the learner's possible misunderstanding.

10.3 Future work

10.3.1 Enhancing the efficiency of the authoring question templates

Even though, in its current implementation, the efficiency of formulating the format of questions can only offer around 70% of selected questions, the architecture of the COMBA system has provided a basic framework that can easily be extended for improving the selected questions.

- **Improving the mechanism of querying data with Natural Language Processing**

Natural Language Processing (NLP) deals with the computer's processing of natural language (language normally used by humans), so that machines can derive semantic understanding of the content.

In the COMBA system, the use of the authoring question template to generate questions was implemented. In order to improve the precise format of questions, the SPARQL queries will have to extract a suitable result set for formulating the questions. Such an improved query mechanism could require adopting NLP techniques to provide a machine learning approach with a rule induction program to learn information extraction rules from subject matter and capability ontology and associated templates.

The system would first extract terminology from the ontology. This approach may rely on machine learning and automated language-processing techniques to extract concepts and ontological relations from the ontology. It would then filter the terminology using NLP and formulate the questions based on semantic interpretation.

- **Enhancing the format of questions**

In the COMBA system, only four simple question templates were implemented. There are some questions that the experts would have expected such as the "Wh" questions. For example, "What is the effect of sample size on the width of a confidence interval?" and "In computing a confidence interval, when do you use 't' and when do you use 'z'?" These were called meta-questions because the topics found in these questions were not extracted directly from the existing ontologies. In order to formulate these questions, more details of ontology would have to be represented by using OWL DL or OWL Full to improve the description of the subject matter hierarchy.

In order to increase the variety of questions, more complicated question templates could require adopting more detailed attributes in these templates.

10.3.2 Extracting ontological vocabulary with NLP

While this study successfully demonstrates a data model and a method of automatically generating acceptable and useful questions, the critical challenge in practice is the representation of competencies and subject matter. Successful deployment of the system would require the development of a detailed and systematic database comprising all the competencies involved in the particular domain of interest.

In the COMBA system, a domain expert expressed domain content in an English-like form. Then, a knowledge engineer represented elements of content with an ontology. This uses a specialist or subject matter content expert to analyse the domain before a knowledge engineer can process it later.

This ontology construction enabled manual ontology engineering to be performed using Protégé development toolsets, including editors, consistency checkers, mediators to support decisions, and ontology import tools. It would be misleading to represent knowledge of the expert with the ontology because the engineer may not be able to understand and represent it as precisely as the expert. This approach could be improved in future by experts expressing their knowledge in the form of ontology themselves.

In addition, a major challenge in the construction of a competency ontology is that the existing competencies in the course syllabus are required to be well-defined and properly represented in the model. This is usually not the case in most existing syllabi, and so future work would need to make explicit connections between course objective, domain content, and capability. Such connections would lead to clear relationships between course objectives, learner assessment, and evaluations of teaching effectiveness.

Furthermore, issues of ontology construction, identifying, defining, and entering concept definitions are critical, because experts can have different points of view about the same concept. This task can be lengthy, costly, and controversial. Future work could develop an alternative approach of defining ontology vocabulary which relies on machine learning and automated language-processing techniques. Such an approach would extract concepts and ontological relations from structured and unstructured data such as databases and text. Automated ontology learning from domain text would then lead to precise representation and suitably formulated questions.

10.3.3 Adaptive instruction using a competency tree

Structuring subject components within this model yields a competency tree where instructional and question sequences could be derived in the same way as the adaptive assessment. Future work would then focus on algorithms for generating pedagogically useful instructional sequences.

Teachers use the competency tree to examine current learning outcomes in units of learning and teaching, and to revise the intended learning outcomes, so that they will align with one another, and with the unit's assessments. Referring to the competency tree gives teachers a place to start when revising units to better align with new standards-based requirements. Future work would focus on integration and interoperability of competency trees from a variety of sources such as education, professional accreditations, and human resource business needs.

10.3.4 Generating feedback for formative assessment

One of main challenges in formative assessment is the creating effective feedback. Effective feedback needs to deliver information that helps learners self-correct and helps clarify what good performance is. A competency tree would allow the generation of feedback derived from nodes semantically close to 'incorrect' nodes in the tree. Feedback could relate to the concepts of the incorrect answer, as well maintain consistence with the key concept of the question. Generating such feedback could be based on pedagogically-driven processes by adapting the traversal algorithm. For example, feedback could be generated from the closest node to incorrect node, requiring the learner to reflect on their answers.

Future work could focus on automatically generating feedbacks which reflect levels of content taxonomy and capability taxonomy, encourage interaction and dialogue around learning, and support self-assessment and reflection in learning. This would allow learners to take more control of their learning and develop their reflective skills.

10.4 Summary

The contribution of this study will enable development of adaptive sequences, computational descriptions, and analyses of pedagogic methods from an ontology database. This Semantic Web technology promises that machines will make more sense of the web content so that intelligent reasoning can be accomplished; hence, potential learning resources can be more accurately discovered.

The development of the competency model has opened up many integrating research outcomes such as adaptive assessment systems, Natural Language Processing and adaptive instruction.

References

AKT Manifesto (2000) http://www.aktors.org/publications/, Accessed. May 15, 2009.

Anderson, T. and Whitelock, D. (2004), The Educational Semantic Web: Visioning and Practicing the Future of Education. *Journal of Interactive Media in Education*, vol. 1, p. 9. http://www-jime.open.ac.uk/2004/1/editorial-2004-1.pdf.

Antoniou, G. and Harmelen, F. v. (2004) *A Semantic Web Primer*: The MIT Press.

Aroyo, L. and Dicheva, D. (2001), AIMS: Learning and Teching Support for WWW-based Education. *International Journal for Continuing Engineering Education and Life-Long Learning*, vol. 11, 1/2, pp. 152-164.

Aroyo, L. and Dicheva, D. (2004), The New Challenges for E-learning: The Educational Semantic Web. *Education Technology and Society*, vol. 7, 4, pp. 59-69.

Askins, D. (2004) Adaptive Assessment *Toolkit - The Nuts and Bolts e-Newsletter*, vol. 3, no. 2.

Atanas, K., Damyan, O., and ManovDimitar OWLIM - : A pragmatic semantic repository for OWL, in *Lecture notes in computer science* I, vol., 2005.

Bailey, C. P. (2001) Using Contextual Information with Dynamic Linkbases to Provide Adaptive Hypermedia, School of Electronics and Computer Science, University of Southampton, Southampton.

Baruque, L. B. and Melo, R. N. (2003), Learning Theory and Instructinal Design Using Learninng Object. *Learning Objects 2003 Symposium: Lessons Learned, Questions Asked.* http://www.cs.kuleuven.ac.be/~erikd/PRES/2003/LO2003/Baruque.pdf

Berger, C. and Kam, R. (1996) Definitions of Instructional Design. Accessed. November 4, 2008.

Berners-Lee, T. (2000) Semantic Web. http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html, Accessed. October 27, 2006.

Berners-Lee, T., Hendler, J., and Lassila, O. (2001) The Semantic Web, *Scientific American*, vol. 284, no. 5. pp. 34-43. http://www.sciam.com/article.cfm?id=the-semantic-web.

Bizer, C., Heese, R., Mochol, M., Oldakowski, R., Tolksdorf, R., and Eckstein, R. The Impact of Semantic Web Technologies on Job Recruitment Processes, in *Wirtschaftsinformatik 2005*I, vol.: Physica-Verlag HD, 2005.

Bloom, B. S. and Krathwohl, D. R. (1956) *Taxonomy of educational objectives: The classification of educational goals by a committee of college and university examiners*. New York: Longman.

Brusilovsky, P. and Pesin, L. (1994) ISIS-Tutor: An adaptive hypertext learning environment. In JCKBSE'94, Russia. pp. 83-87.

Brusilovsky, P. (1996), Methods and Techniques of Adaptive Hypermedia. *User Modeling and User-Adapted Interaction*, vol. 6, 2-3, pp. 87-129.

Brusilovsky, P. (2001), Adaptive Hypermedia. *User Modeling and User-Adapted Interaction*, vol. 11, 1-2, pp. 87-100.

Brusilovsky, P. and Sosnovsky, S. (2005), Individualized exercises for self-assessment of programming knowledge: An evaluation of QuizPACK. *Educational Resources in Computing (JERIC)*, vol. 5, 3, p. 6.

Burgos, D. and Griffiths, D. (2005) Understanding and using Learning Design, in *The UNFOLD Project*, 2005.

Cheetham, G. and Chivers, G. (2006), Professions, competence and informal learning. *International Journal of Training and Development*, vol. 10, 2, p. 175.

Cheniti-Belcadhi, L. and Braham, R. (2004) A generic framework for assessment in adaptive educational hypermedia. In Proceedings of IADIS International Conference WWW/Internet 2004 (ICWI 2004), Madrid, Spain. http://www.kbs.uni-hannover.de/Arbeiten/Publikationen/2004/iadis04.pdf.

Cheniti-Belcadhi, L., Henze, N., and Braham, R. (2005) Towards a Service Based Architecture for Assessment. In Proceedings of 13th Annual Workshop of the SIG Adaptivity and User Modeling in Interactive Systems (ABIS 2005), Saarbrücken, Germany. http://rewerse.net/publications/download/REWERSE-RP-2005-108.pdf.

Chew, L. K. (2001), Use of XML in eLearning Content Packaging. www.itsc.org.sg/synthesis/2001/itsc-synthesis2001-kinchew-xml-eLearning.pdf

Collins, J. A., Greer, J. E., and Huang, S. X. (1996), Adaptive assessment of using granularity hierarchies and Bayesien nets. *Proceedings of Intelligent Tutoring Systems*, pp. 569-577. http://citeseer.ist.psu.edu/collins96adaptive.html

Conejo, R., Guzmán, E., Eva Millán, M. T., Pérez-De-La-Cruz, J. L., and Ríos, A. (2004), SIETTE: A Web–Based Tool for Adaptive Testing. *International Journal of Artificial Intelligence in Education*, vol. 1, 33. http://aied.inf.ed.ac.uk/members04/archive/Vol_14/Conejo/Conejo04.pdf.

Conole, G. and Fill, K. (2005), A learning design toolkit to create pedagogically effective learning activities. *Journal of Interactive Media in Education 2005*. http://jime.open.ac.uk/2005/08/

CopperCore (2005) http://coppercore.sourceforge.net/ Accessed. November 4, 2008.

Cristea, P. and Tuduce, R. (2005) Automatic Generation of Exercises for Self-testing in Adaptive E-learning Systems: Exercises on AC Circuits, in *3rd Workshop on Adaptive and Adaptable Educational Hypermedia at the AIED'05 conference, (A3EH)*, 2005.

Dancik, G. and Kumar, A. (2003) A Tutor for Counter-Controlled Loop Concepts and Its Evaluation. In Frontiers in Education Conference, USA.

Davis, H. C. (1999) Rapid Construction of Learning Resources using Microcosm Pro, *Connections*, vol. 11, no. 1. pp. 14-17.

Davis, H. C. and White, S. (2001), Linking Experiences: Issues Raised Developing Linkservices for Resource Based Learning and Teaching. *In Proceedings of IEEE International Conference on Advanced Learning Technologies*.

Davis, R., Schrobe, H., and Szolovits, P. (1993) What is knowledge representation?, *AI Magazine*, vol. 14, no. 1. pp. 17-33.

De Bra, P. (2000), Pros and cons of adaptive hypermedia in Web-based education. *CyperPsychology and Behavior*, vol. 3, 1, pp. 71-77.

De Bra, P., Aerts, A., Berden, B., Lange, B. d., Rousseau, B., Santic, T., Smits, D., and Stash, N. (2003), AHA! The adaptive hypermedia architecture. *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*.

De Bra, P., Aroyo, L., and Chepegin, V. (2004), The Next Big Thing: Adaptive Web-Based Systems. *Journal of Digital Information*, vol. 5, 1.

Dicheva, D., Dichev, C., and Wang, D. (2005) Visualizing Topic Maps for e-Learning. In Proceedings of IEEE International Conference on Advanced Learning Technologies, Kaohsiung, Taiwan. http://ieeexplore.ieee.org/iel5/10084/32317/01508866.pdf.

Dicheva, D., Sosnovsky, S., Gavrilova, T., and Brusilovsky, P. (2005) Ontological Web Portal for Educational Ontologies. In Artificial Intelligence in Education, Amsterdam, Netherlands.

Dick, W. and Cary, L. (1990) *The Systematic Design of Instruction*: Harper Collins.

Dick, W. and Cary, L. (1996) *The Systematic Design of Instruction*. New York: Haper Collins College Publishers.

Draganidis, F. and Mentzas, G. (2006), Competency based management: a review of systems and approaches *Information Management & Computer Security*, vol. 14, 1, pp. 51-64.

E-Learning Engineering (2004) The Challenge of E-Learning Development. http://www.elearning-engineering.com/learning/e-learning/engineering-elearning.htm Accessed. November 4, 2008.

Ertmer, P. A. and Newby, T. J. (1993), Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, vol. 6, 4, pp. 50-70.

Falmagne, J.-C., Cosyn, E., Doignon, J.-P., and Thiery, N. (2003) The Assessment of Knowledge, in Theory and in Practice. In Integration of Knowledge Intensive Multi-Agent Systems. pp. 609-615. http://www.aleks.com/about_aleks/Science_Behind_ALEKS.pdf.

Faul, F., Erdfelder, E., Lang, A.-G., and Bucnher, A. (2007), G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, vol. 2, 39, pp. 175-191.

Field, A. (2005) *Discovering Statistics Using SPSS*. London: SAGE Publications Ltd.

Friensen, N. and Anderson, T. (2004), Interaction for lifelong learning. *British Journal of Educational Technology*, vol. 35, 6, pp. 679-687.

Gadomski, A. M. (2001) Knowledge structure, 2001. http://erg4146.casaccia.enea.it/K-struct.htm.

Gagne, R. (1970) Robert Gagne's Nine Steps of Instruction. http://www.nwlink.com/~donclark/hrd/learning/development.html Accessed. January 3, 2006.

Gagne, R. M. The Conditions of Learning and the Theory of Instruction,New York: Holt, Rinehart, and Winston., 2006I, 1985.

Gennari, J. H., Musen, M. A., Fergerson, R. W., Grosso, W. E., Crubézy, M., Eriksson, H., Noy, N. F., and Tu, S. W. (2003), The Evolution of Protégé: An Environment for Knowledge-Based Systems Development. *International Journal of Human-Computer Studies*, vol. 58, 1, pp. 89-123.

Gilbert, L., Sim, Y. W., and Wang, C. (2005a), Modelling the Learning Transaction. *In Proceedings of The 5th IEEE International Conference on Advanced Learning Technologies*.

Gilbert, L., Sim, Y. W., and Wang, C. (2005b), An e-Learning Systems Engineering Methodology. *In Proceedings of The 5th IEEE International Conference on Advanced Learning Technologies*. http://eprints.ecs.soton.ac.uk/10650/

Gilbert, L. and Gale, V. (2007) *Principles of eLearning Systems Engineering*: Chandos.

Gogoulou, A., Gouli, E., Grigoriadou, M., and Samarakou, M. (2004), Adapting the Communication-Scaffolding Tools in a Web-based Collaborative Learning Environment. *Proceedings of the ED-MEDIA 2004*, vol. 2004, 1, pp. 1153-1161.

Gómez-Pérez, A. and Manzano-Macho, D. (2004), An overview of methods and tools for ontology learning from texts. *The Knowledge Engineering Review*, vol. 19, 3, pp. 187-212.

Gonzalez, A. J. and Dankel, D. D. (1993) *The Engineering of Knowledge-based Systems Theory and Practice*: Prentice Hall.

Gouli, E., Gogoulou, A., Papanikolaou, K., and Grigoriadou, M. (2004) COMPASS: An adaptive web-based concept map assessment toool. In First Int. Conference on Concept Mapping, Pamplona, Spain. http://cmc.ihmc.us/papers/cmc2004-128.pdf#search=%22COMPASS%3A%20An%20Adaptive%20web-based%20COncept%20Mapping%22.

Griffiths, D. and Blat, J. (2005), The role of teachers in editing and authoring units of learning using IMS Learning Design. *Advanced Technology for Learning*, vol. 2, 4. http://dspace.ou.nl/bitstream/1820/586/1/griffiths_atl_2005.pdf

Gruber, T. (1993), Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies*, vol. 43, 5-6, pp. 907-928.

Hersh, W. R., Bhupatiraju, R. T., Greene, P. S., Smothers, V., and Cohen, C. (2006) Adopting e-learning standards in health care: competency-based learning in the medical informatics domain. In AMIA Annual Symposium Proceedings 2006. pp. 334-338.

Horton, W. K. (2006) *E-Learning by Design*. San Francisco, USA: Pfeiffer An Imprint of Wiley.

HR-XML Consortium (2006) Competencies (Measurable Characteristics) Recommendation. http://ns.hr-xml.org/2_4/HR-XML-2_4/CPO/Competencies.html, Accessed. November 4, 2008.

IMS Content Packaging (2005) IMS Content Packaging Overview http://www.imsglobal.org/content/packaging/cpv1p2pd/imscp_oviewv1p2pd.html Accessed. November 4, 2008.

IMS LD (2003) IMS Learning Design Best Practice and Implementation Guide http://www.imsglobal.org/learningdesign/ldv1p0/imsld_bestv1p0.html, Accessed. November 4, 2008.

IMS QTI (2006) IMS Question and Test Interoperability Overview. http://www.imsglobal.org/question/qtiv2p1pd2/imsqti_oviewv2p1pd2.html Accessed. November 4, 2008.

IMS RDCEO (2002) IMS Resuable Definition of Competency or Educational Objective-Best Practice, 1.0 Final Specification. http://www.imsglobal.org/competencies/index.html, Accessed. November 4, 2008.

JISC (2004) Effective Practice with e-Learning: JISC, 2004.

Johnson, B. and Shneiderman, B. (1991) Tree-Maps: a space-filling approach to the visualization of hierarchical information structures In Proceedings of the 2nd conference on Visualization '91 San Diego, California pp. 284-291.

Joint SFEFC/SHEFC E-Learning Group (2003) *Final Report*.

Jovanovic, J., Gasevic, D., and Devedzic, V. (2006), Ontology-Based Automatic Annotation of Learning Content. *Semantic Web and Information Systems*, vol. 2, 2, pp. 91-119.

Kalfoglou, Y. (2001) *Exploring Ontologies*. vol. 1: Fundamentals, Handbook of Software Engineering and Knowledge Engineering. Washington, D.C.: World Scientific Publishing.

Karampiperis, P., Sampson, D., and Fytros, D. (2006) Lifelong Competence Development: Towards a Common Metadata Model for Competencies Description - The Case Study of Europass Language Passport. In Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies, Kerkrade, The Netherlands. http://dspace.learningnetworks.org/bitstream/1820/683/1/312karsamfyt.pdf.

Kemp, J. E., Morrison, G. R., and Ross, S. M. (1998) *Designing Effective Instruction*. vol. Prentice Hall, Second Edition ed. Ohio.

Kolås, L. and Staupe, A. (2004) Implementing delivery methods by using pedagogical design patterns. In World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA) 2004. pp. 5304-5309. http://dl.aace.org/16250

Kommers, P., Grabinger, S., and Dunlap, J. (1996), Hypermedia Learning Environments. *Instructional Design and Integration*.

Krathwohl, D. R. and Anderson, L. (2002), A revision of bloom's taxonomy: An overview. *Theory into Practice*, vol. 41, 4, pp. 212-218.

Kunzmann, C. (2006) Ontology-based Competence Management for Healthcare Training Planning: A Case Study. In Proceeding of the International Conference on Knowledge Management, Austria.

Lalos, P., Retalis, S., and Psaromiligkos, Y. (2005), Creating personalised quizzes both to the learner and to the access device characteristics: the Case of CosyQTI. http://wwwis.win.tue.nl/~acristea/AAAEH05/papers/3-camera%20ready-CoSyQTI.pdf.

Laurillard, D. (2002) *Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies*, 2nd edition ed. London: RoutledgeFalmer.

Lukose, D. (1996) *Knowledge Engineering*. Australia: University of New England.

Mayes, T. and Freitas, S. d. (2004) Stage2: Review of e-learning theories, frameworks and models: JISC, 2004.

McComas, W. F. and Abraham, L. (2005) Asking more effective questions. http://www.usc.edu/programs/cet/private/pdfs/usc/Asking_Better_Questions.pdf, Accessed. October 27, 2007.

Mergel, B. (1998) Instructional Design and Learning Theory. http://www.usask.ca/education/coursework/802papers/mergel/brenda.htm Accessed. November 4, 2008.

Merrill, M. D. (1983) Component Display Theory. http://coe.sdsu.edu/eet/articles/cdt/index.htm, Accessed. November 4, 2008.

Merrill, M. D. (1999) Instructional Transaction Theory (ITT): Instructional Design Based on Knowledge Objects. http://www.indiana.edu/~idtheory/chapter_17_summary.html Accessed. January 3, 2006.

Merrill, M. D. (2002), A Pebble-in-the-Pond Model for Instructional Design. *Performance Improvement*, vol. 41, 7, pp. 39-44.

Moreau, L., Gibbins, N., DeRoure, D., El-Beltagy, S., Hall, W., Hughes, G., Joyce, D., Kim, S., Michaelides, D., Millard, D., Reich, S., Tansley, R., and Weal, M. (2000) SoFAR with DIM Agents: An Agent Framework for Distributed Information Management. In he Fifth International Conference and Exhibition on The Practical Application of Intelligent Agents and Multi-Agents.

Myller, N. (2007), Automatic Generation of Prediction Questions during Program Visualization. *Electronic Notes in Theoretical Computer Science (ENTCS)*, vol. 178, pp. 43-49.

Noy, N. F. and McGuinness, D. L. (2001) Ontology Development 101: A Guide to Creating Your First Ontology, Stanford University 2001.

Papanikolaou, K. A., Grigoriadou, M., Magoulas, G. D., and Kornilakis, H. (2002), Towards New Forms of Knowledge Communication: the Adaptive Dimension of a Web-based Learning Environment. vol. 39, 4, pp. 333-360.

Papanikolaou, K. A. and Grigoriadou, M. (2003), An Instructional Framework Supporting Personalized Learning on the Web. *The 3rd IEEE International Conference on Advanced Learning Technologies*.

Paquette, G. (2007), An Ontology and a Software Framework for Competency Modeling and Management. *Educational Technology & Society*, vol. 10, 3, pp. 1-21.

Prakash, P. R., Sasikumar, M., Mhashilkar, P., and Vakil, R. (2000) Vyasa: A system for multiple-choice generative questions. In International conference on Knowledge Based Computer Systems, Mumbai, India.

Reich, S. and Davis, H. C. (1999), The need for an open hypertext protocol. *ACM SIGWeb Newsletter* vol. 8, 1, pp. 21-23.

Reigeluth, C. M. (1999), Instructional Design Theories and Models: A New Paradigm of Instructional Theory. vol. 2.

Reload (2004) Reload editor. http://www.reload.ac.uk/ Accessed. November 4, 2008.

Reload (2006) Reload Learning Design editor. http://www.reload.ac.uk/ldeditor.html Accessed. November 4, 2008.

Rogers, C. (1983) *Freedom to Learn for the 80s*. New York: Merrill.

Ryder, M. (2006) Instructional Design Models. http://carbon.cudenver.edu/~mryder/itc_data/idmodels.html Accessed. November 4, 2008.

Sandberg, R. Competence-the Basis for a Smart Workforce, R. Gerber and C. Lankshear (Eds.) Training for a smart workforce, Ed. London: Routledge, 2000, pp. 47-72.

Schmidt, A. and Kunzmann, C. (2006) Towards a Human Resource Development Ontology for Combining Competence Management and Technology-Enhanced Workplace Learning. In Proceedings of OntoContent 2006 pp. 1078-1087.

SCORM (2004) SCORM 2004 Overview. http://www.adlnet.gov/news/articles/375.cfm Accessed. November 4, 2008.

Silyn-Roberts, H. (1996) *Writing for Science*. Auckland, New Zealand: Addison Wesley Longman New Zealand Limited.

Sitthisak, O., Gilbert, L., and Davis, H. C. (2007) Towards a competency model for adaptive assessment to support lifelong learning. In TENCompetence Workshop on Service Oriented Approaches and Lifelong Competence Development Infrastructures, Manchester, UK.

Sitthisak, O., Gilbert, L., Davis, H. C., and Gobbi, M. (2007a) Adapting health care competencies to a formal competency model. In the ICALT, Niigata, Japan.

Sitthisak, O., Gilbert, L., Zalfan, M. T., and Davis, H. C. (2007b) Interactivity within IMS Learning Design and Question and Test Interoperability. In Proceedings of Proceedings of the 3rd International Conference on Web Information Systems and Technologies (WEBIST), Barcelona, Spain.

Sitthisak, O., Gilbert, L., and Davis, H. C. (2008a) Deriving e-assessment from a competency model. In The 8th IEEE International Conference on Advanced Learning Technologies (ICALT 2008), Santander, Cantabria, Spain.

Sitthisak, O., Gilbert, L., and Davis, H. C. (2008b), An evaluation of pedagogically informed parameterised questions for self assessment. *Learning, Media and Technology*, vol. 33, 3, pp. 235-248.

Sitthisak, O., Gilbert, L., and Davis, H. C. (2008c) Transforming a competency model to assessment items. In The 4th International Conference on Web Information Systems and Technologies (WEBIST), Funchal, Madeira - Portugal.

Sitthisak, O., Gilbert, L., and Davis, H. C. (2009) *Transforming a Competency Model to Parameterised Questions in Assessment*. vol. 18, Lecture Notes in Business Information Processing: Springer-Verlag Berlin Heidelberg.

Smith, P. (1996) *An Introduction to Knowledge Engineering*: International Thompson Computer Press.

Smith, P. L. and Ragan, T. J. (1999) *Instructional design*. New York: John Wiley & Sons.

Sowa, J. F. (2000) *Knowledge representation: logical, philosophical and computational foundations*: Brooks/Cole Publishing Co.

Stojanovic, L., Staab, S., and Studer, R. (2001) eLearning based on the Semantic Web. In World Conference on the WWW and Internet, Orlando, Florida, USA. pp. 23-27. http://www.aifb.uni-karlsruhe.de/WBS/Publ/2001/WebNet_lstsstrst_2001.pdf.

The European Commission (2006) Lifelong learning. http://ec.europa.eu/education/policies/lll/life/what_islll_en.html, Accessed. November 4, 2008.

Trichet, F. and Leclère, M. (2003) *A Framework for Building Competency-Based Systems Dedicated to Human Resource Management*. vol. 2871/2003, Lecture Notes in Computer Science: Springer Berlin / Heidelberg.

Tzanavari, A., Retalis, S., and Pastellis, P. (2004) Giving More Adaptation Flexibility to Authors of Adaptive Assessments. In Adaptive Hypermedia 2004 Conference Proceedings. http://www.softlab.ntua.gr/~retal/papers/VOLUMES/springer_LMS/ah2004-AdaptiveTesting.pdf#search=%22Giving%20More%20Adaptation%20Flexibility%20to%20Authors%20of%20Adaptive%20Assessment%22.

W3C (2004a) OWL Web Ontology Language (OWL). http://www.w3.org/TR/owl-ref/, Accessed. October 27, 2007.

W3C (2004b) OWL Web Ontology Language Use Cases and Requirements. http://www.w3.org/TR/webont-req/#onto-def, Accessed. October 27, 2007.

W3C (2005) SKOS Core Guide. http://www.w3.org/TR/swbp-skos-core-guide/, Accessed. November 4, 2008.

Way, W. D. (2005) Practical Questions in Introducing Computerized Adaptive Testing for K-12 Assessments, 2005.

Weaver, P. L. (1993) *Practical SSADM*. Pitman, London.

Weber, G. and Brusilovsky, P. (2001), ELM-ART: An adaptive versatile system for Web-based instruction. *International Journal of Artificial Intelligence in Education Special Issue on Adaptive and Intelligent Web-based Educational Systems*, vol. 12, 4, pp. 351-384.

Welch, R. E. and Frick, T. W. (1993), Computerized adaptive testing in instructional settings. *Educational Technology Research and Development*, vol. 41, 3, pp. 47-62.

Wills, G., Davis, H., Chennupati, S., Gilbert, L., Howard, Y., Jam, E.-R., Jeyes, S., Millard, D., Sherratt, R., and Willingham, G. (2006) R2Q2: Rendering and Reponses Processing for QTIv2 Question Types. In 10th International CAA Conference, Loughborough University, UK.

Woukeu, A., Wills, G., Conole, G., Carr, L., Kampa, S., and Hall, W. (2003) Ontological Hypermedia in Education: A framework for building web-based educational portals. In World Conference on Educational Multimedia, Hypermedia & Telecommunications, Honolulu, Hawaii, USA.

XIA Systems Corporations (2005) Rapid eLearning Development Process. http://www.xiasystems.com/cSolutions/eProcess.htm Accessed. November 4, 2008.

Appendix A – Example of QTI question in XML format

```
  <?xml version="1.0" encoding="UTF-8" ?>
- <assessmentItem
    xsi:schemaLocation="http://www.imsglobal.org/xsd/imsqti_v2p0
    imsqti_v2p0.xsd" title="Question" adaptive="false" timeDependent="false"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    identifier="question20" xmlns="http://www.imsglobal.org/xsd/imsqti_v2p0">
  + <responseDeclaration baseType="identifier" cardinality="single"
      identifier="RESPONSE">
  + <responseDeclaration baseType="integer" cardinality="single"
      identifier="sequencecap">
  + <responseDeclaration baseType="integer" cardinality="single"
      identifier="sequenceSMC">
  + <responseDeclaration baseType="integer" cardinality="single"
      identifier="subSequence">
  + <responseDeclaration baseType="integer" cardinality="single"
      identifier="subSequenceSMC">
  + <outcomeDeclaration baseType="integer" cardinality="single" identifier="SCORE">
    <outcomeDeclaration identifier="FEEDBACK" cardinality="single"
      baseType="identifier" />
  - <itemBody>
      <p />
      <p />
    - <choiceInteraction responseIdentifier="RESPONSE" shuffle="false"
        maxChoices="1">
      - <p>
        - <b>
            <prompt>The formula of the complexity adjustment factor
              may be defined as</prompt>
          </b>
        </p>
        <simpleChoice identifier="Choice2">The aggregate of the ratings of the
          19 influence factors.</simpleChoice>
        <simpleChoice identifier="Choice1">0.65 plus degree of influence
          divided by 200</simpleChoice>
        <simpleChoice
          identifier="Choice3">0.26*tof+0.56*tif+1.66*tea</simpleChoice>
        <simpleChoice identifier="Choice4">200 plus 0.65 divided by degree of
          influence</simpleChoice>
      </choiceInteraction>
    </itemBody>
  + <responseProcessing>
    <modalFeedback outcomeIdentifier="FEEDBACK" identifier="Choice1"
      showHide="show">Yes, that is correct.</modalFeedback>
    <modalFeedback outcomeIdentifier="FEEDBACK" identifier="Choice1"
      showHide="hide">No, the correct answer is "0.65 plus degree of influence
      divided by 200".</modalFeedback>
  </assessmentItem>
```

Appendix B – Experiment 1 questionnaire

Questionnaire 1

A variety of self-assessment questions have been prepared to help you develop and assess your proficiencies. These self-assessment questions differ in their wording and content in a number of ways, and we would appreciate your opinions on them.

Please read the self-assessment questions and then evaluate them by selecting the option which most closely reflects your opinion. Please do not attempt to answer any of the self-evaluation questions, and please do not spend too much time considering them. It is your first impression which is most valuable.

PART 1

This questionnaire asks you to rate a number of questions against four criteria for self-assessment. The criteria are:

Clear means the question is easy to understand.

Match means the question relates to the learning outcome, "Students understand the concept of a confidence interval, and can calculate it."

Useful means the question is useful for further study and your own self-assessment.

Challenging means the question is appropriately difficult.

Please tick each criterion that applies if it describes your opinion of the question.

Question	Clear	Match	Useful	Challenging
Calculate the confidence interval for this situation: Nine hundred (900) high school first year students were randomly selected for a national survey. Among survey participants, the mean grade-point average (GPA) was 2.7, and the population standard deviation was 0.4 assuming a 95% confidence level.				

Question	Clear	Match	Useful	Challenging
Define the meaning of the confidence interval.				

Question	Clear	Match	Useful	Challenging
Explain the importance of the confidence interval.				

Question	Clear	Match	Useful	Challenging
Define the meaning of the confidence interval for this situation: Nine hundred (900) high school first year students were randomly selected for a national survey. Among survey participants, the mean grade-point average (GPA) was 2.7, and the population standard deviation was 0.4 assuming a 95% confidence level.				
Question	**Clear**	**Match**	**Useful**	**Challenging**
Explain the importance of a confidence interval for this situation: Nine hundred (900) high school first year students were randomly selected for a national survey. Among survey participants, the mean grade-point average (GPA) was 2.7, and the population standard deviation was 0.4 assuming a 95% confidence level.				

Please feel free to write any further comments which you feel would improve in the development of the questions.

PART 2

When did you first study statistics?

O for A level O at University O Other ________

Appendix C – Experiment 2 questionnaire

Questionnaire 2

Please use the following instruction to test the system.

Please answer the following question by selecting the option which most closely reflects your opinion.

PART 1: Background Information

1.1 **Do you know what an adaptive assessment system is?**

□ Yes □ No

1.2 **Have you ever used a system or website that can adapt depending upon your prior selection?**

□ Yes □ No

If "yes" Do you like this approach? □ Yes □ No □ N/A

1.3 **For self-assessment, have you ever tried asking yourself questions and attempting to answer them?**

□ Yes □ No

If "yes" Do you like this approach? □ Yes □ No □ N/A

PART 2: Testing the system

Look at the website "**http://playr.qtitools.org/".**

Please select the assessment "**Function Point Analysis 1**" from the drop down list.

Then, click "**Start Assessment**" button.

Next, answer the questions until the end of the test. **Please do not skip any answers.**

2.1 **How many presented questions did you get? _____________**

2.2 **What was your percentage of correct answers? _____________**

PART 3: Evaluating the test

Please tick the option that best describes your opinion on this statement, when you finished the test.

3.1 **The test fairly assesses the level of my knowledge of the following learning outcomes.**

"Main Learning outcomes:

Student will be able to calculate Adjusted Function Points with some situation."

"Sub Learning outcomes:

"1 Student will be able to calculate Complexity Adjustment."

"2 Student will be able to calculate Unadjusted Function Points."

□ Strongly Agree □ Agree □ Disagree □ Strongly disagree

PART 4: Evaluating the adaptation of the question

4.1 Look at the website "**http://playr.qtitools.org/".**

Please select the assessment "**Function Point Analysis 2**" from the drop down list.

Then, click "**Start Assessment**" button.

Next, answer **the first question** with **the first choice.**

Then**,** click **"Submit answer"** and **"Next"** button.

Considering **the next question**,

4.1.1 **What is the next question after selecting the first choice? (Please write just the first line and Question Number)**

Then, click "**Exit test**" button and "**Yes**".

4.2 Look at the website "**http://playr.qtitools.org/".**

Please select an assessment which is "**Function Point Analysis 2**" from the drop down list.

Then, click "**Start Assessment**" button.

Next, answer **the first question** with **the second choice. (The correct answer)**

Then**,** click **"Submit answer"** and **"Next"** button.

Considering **the next question**,

4.2.1 **What is the next question after selecting the correct answer? (Please write just the first line and Question Number)**

Please tick the option that best describes your opinion on this statement.

4.3 **The system adapted the next question based on my answer.**

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly disagree

4.4 **It was useful for self-assessment, if the test adapt the next question based on my answer.**

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly disagree

4.5 **The system helped me to identify what I do not know based on a given learning outcome?**

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly disagree

4.6 **The system helped me to understand how a given learning outcome breaks down into learning outcome components.**

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly disagree

4.7 **The system helped me to understand how a given learning outcome breaks down into topics.**

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly disagree

4.8 **The system showed me an appropriately difficult question based on my previous answer.**

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly disagree

PART 5: Evaluating the question

Please tick the option that best describes your opinion on this statement.

5.1 **The questions in the test were clear.**

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly disagree

5.2 **The questions in the test fairly assessed the level of my knowledge of the following learning outcomes.**

"Main Learning outcomes:

Student will be able to calculate Adjusted Function Points with some situation."

"Sub Learning outcomes:

"1 Student will be able to calculate Complexity Adjustment."

"2 Student will be able to calculate Unadjusted Function Points."

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly disagree

5.3 **The questions in the test were useful for further study.**

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly disagree

5.4 **The questions in the test were useful for my own self-assessment.**

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly disagree

5.5 **The questions in the test were appropriately difficult.**

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly disagree

Please feel free to write any further comments which you feel would improve in the development of the questions and testing.

Appendix D – Experiment 3.1 questionnaire

Questionnaire 3.1

Please use the following instruction to test the system.

Please answer the following question by selecting the option which most closely reflects your opinion.

PART 1: Background Information

1.1 **Do you know what an adaptive assessment system is?**

 Yes No

1.2 **Have you ever used a system or website that can adapt depending upon your prior selection?**

 Yes No

If "yes" Do you like this approach? Yes No N/A

1.3 **For self assessment, have you ever tried asking yourself questions and attempting to answer them?**

 Yes No

If "yes" Do you like this approach? Yes No N/A

PART 2: Testing the system with the sequence #1

Look at the website http://users.ecs.soton.ac.uk/os05r/info2007.htm

Read the objectives and the scenario of the experiment.

Click "**continue**" button. Then read the given learning outcomes and click "**Begin to generate questions**".

Please select the assessment "**FPA Top down method D**" from the drop down list.

Then, click "**Start Assessment**" button.

Read the question and Select the answer.

Please note: In general, it is not necessary to refer to your lecture notes or to calculate any values in order to answer the questions. We would like you to answer given your current knowledge and understanding.

For the first question, Please select the first or the second or the third answer. (The fourth answer is the correct answer; please do not select that answer.)

Click "**Submit answer**" button. After you have received feedback, click "**Next**" button

Answer the questions until the end of the test. **Please do not skip any answers.**

2.1 **How many presented questions did you get? _____________**

2.2 **What was your percentage of correct answers? _____________**

PART 3: Evaluating the test

Please tick the option that best describes your opinion on this statement, when you finished the test.

3.1 **The test fairly assesses the level of my knowledge of the following learning outcomes.**

"Main Learning outcome:

Student will be able to calculate Adjusted Function Points with some situation."

"**Sub Learning outcomes:**

"1 Student will be able to calculate Complexity Adjustment."

"2 Student will be able to calculate Unadjusted Function Points."

□ Strongly Agree □ Agree □ Disagree □ Strongly disagree

3.2 **How well did the adaptive sequencing help you to identify what you do not know?**

□ Wonderful □ Pretty good □ Acceptable □ Poor □ Totally poor

3.3 **How well did the adaptive sequencing help you to understand how a given learning outcome breaks down into learning outcome components?**

□ Wonderful □ Pretty good □ Acceptable □ Poor □ Totally poor

3.4 **How well did the adaptive sequencing help you to understand how a given learning outcome breaks down into topics?**

□ Wonderful □ Pretty good □ Acceptable □ Poor □ Totally poor

3.5 **How well did the adaptive sequencing provide a complete assessment of the level of your knowledge?**

□ Wonderful □ Pretty good □ Acceptable □ Poor □ Totally poor

3.6 **How useful was the adaptive sequencing for your own self-assessment?**

□ Wonderful □ Pretty good □ Acceptable □ Poor □ Totally poor

PART 4: Testing the system with the sequence#2

4.1 Look at the website "**http://playr.qtitools.org/".**

Please select the assessment "**FPA Bottom up method D**" from the drop down list.

Then, click "**Start Assessment**" button.

Read the question and Select the answer.

How many presented questions did you get? _____________

What was your percentage of correct answers? _____________

4.2 **The test fairly assesses the level of my knowledge of the following learning outcomes.**

"**Main Learning outcome:**

Student will be able to calculate Adjusted Function Points with some situation."

"**Sub Learning outcomes:**

"1 Student will be able to calculate Complexity Adjustment."

"2 Student will be able to calculate Unadjusted Function Points."

□ Strongly Agree □ Agree □ Disagree □ Strongly disagree

4.3 **How well did the adaptive sequencing help you to identify what you do not know?**

□ Wonderful □ Pretty good □ Acceptable □ Poor □ Totally poor

4.4 **How well did the adaptive sequencing help you to understand how a given learning outcome breaks down into learning outcome components?**

□ Wonderful □ Pretty good □ Acceptable □ Poor □ Totally poor

4.5 **How well did the adaptive sequencing help you to understand how a given learning outcome breaks down into topics?**

□ Wonderful □ Pretty good □ Acceptable □ Poor □ Totally poor

4.6 **How well did the adaptive sequencing provide a complete assessment of the level of your knowledge?**

□ Wonderful □ Pretty good □ Acceptable □ Poor □ Totally poor

4.7 **How useful was the adaptive sequencing for your own self-assessment?**

□ Wonderful □ Pretty good □ Acceptable □ Poor □ Totally poor

Please feel free to write any further comments which you feel would improve in the development of the questions and testing.

Appendix E – Tests whose results were not significant

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Order	1.403	1	1.403	2.080	0.158
Algorithm	0.568	1	0.568	0.842	0.365
StartingPoint	0.046	1	0.046	0.069	0.795
Order * Algorithm	0.012	1	0.012	0.017	0.896
Order * StartingPoint	0.742	1	0.742	1.100	0.302
Algorithm * StartingPoint	0.742	1	0.742	1.100	0.302
Order * Algorithm * StartingPoint	0.046	1	0.046	0.069	0.795
Error	22.933	34	0.675		

Table B-1 Tests of Between-Subjects Effects for TestAssessKw

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Order	0.609	1	0.609	0.454	0.505
Algorithm	0.036	1	0.036	0.026	0.872
StartingPoint	0.453	1	0.453	0.337	0.565
Order * Algorithm	0.209	1	0.209	0.156	0.695
Order * StartingPoint	0.453	1	0.453	0.337	0.565
Algorithm * StartingPoint	0.992	1	0.992	0.739	0.396
Order * Algorithm * StartingPoint	0.122	1	0.122	0.091	0.764
Error	45.633	34	1.342		

Table B-2 Tests of Between-Subjects Effects for IdentKw

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Order	1.278	1	1.278	1.390	0.247
Algorithm	0.026	1	0.026	0.028	0.867
StartingPoint	0.186	1	0.186	0.202	0.656
Order * Algorithm	0.235	1	0.235	0.255	0.617
Order * StartingPoint	1.159	1	1.159	1.261	0.269
Algorithm * StartingPoint	0.186	1	0.186	0.202	0.656
Order * Algorithm * StartingPoint	0.742	1	0.742	0.807	0.375
Error	31.267	34	0.920		

Table B-3 Tests of Between-Subjects Effects for DecomposeLearningOutcome

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Order	2.036	1	2.036	2.082	.158
Algorithm	0.209	1	0.209	0.214	0.646
StartingPoint	0.018	1	0.018	0.019	0.893
Order * Algorithm	0.262	1	0.262	0.268	0.608
Order * StartingPoint	0.036	1	0.036	0.036	0.850
Algorithm * StartingPoint	0.609	1	0.609	0.623	0.435
Order * Algorithm * StartingPoint	0.036	1	0.036	0.036	0.850
Error	33.233	34	0.977		

Table B-4 Tests of Between-Subjects Effects for DecomposeTopic

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Order	1.601	1	1.601	1.593	0.215
Algorithm	0.453	1	0.453	0.451	0.507
StartingPoint	0.007	1	0.007	0.006	0.936
Order * Algorithm	0.383	1	0.383	0.381	0.541
Order * StartingPoint	0.320	1	0.320	0.318	0.577
Algorithm * StartingPoint	0.320	1	0.320	0.318	0.577
Order * Algorithm * StartingPoint	0.320	1	0.320	0.318	0.577
Error	34.167	34	1.005		

Table B-5 Tests of Between-Subjects Effects for ProvideCompleteAssessment

Appendix F – Experiment 3.2 questionnaire

Questionnaire 3.2

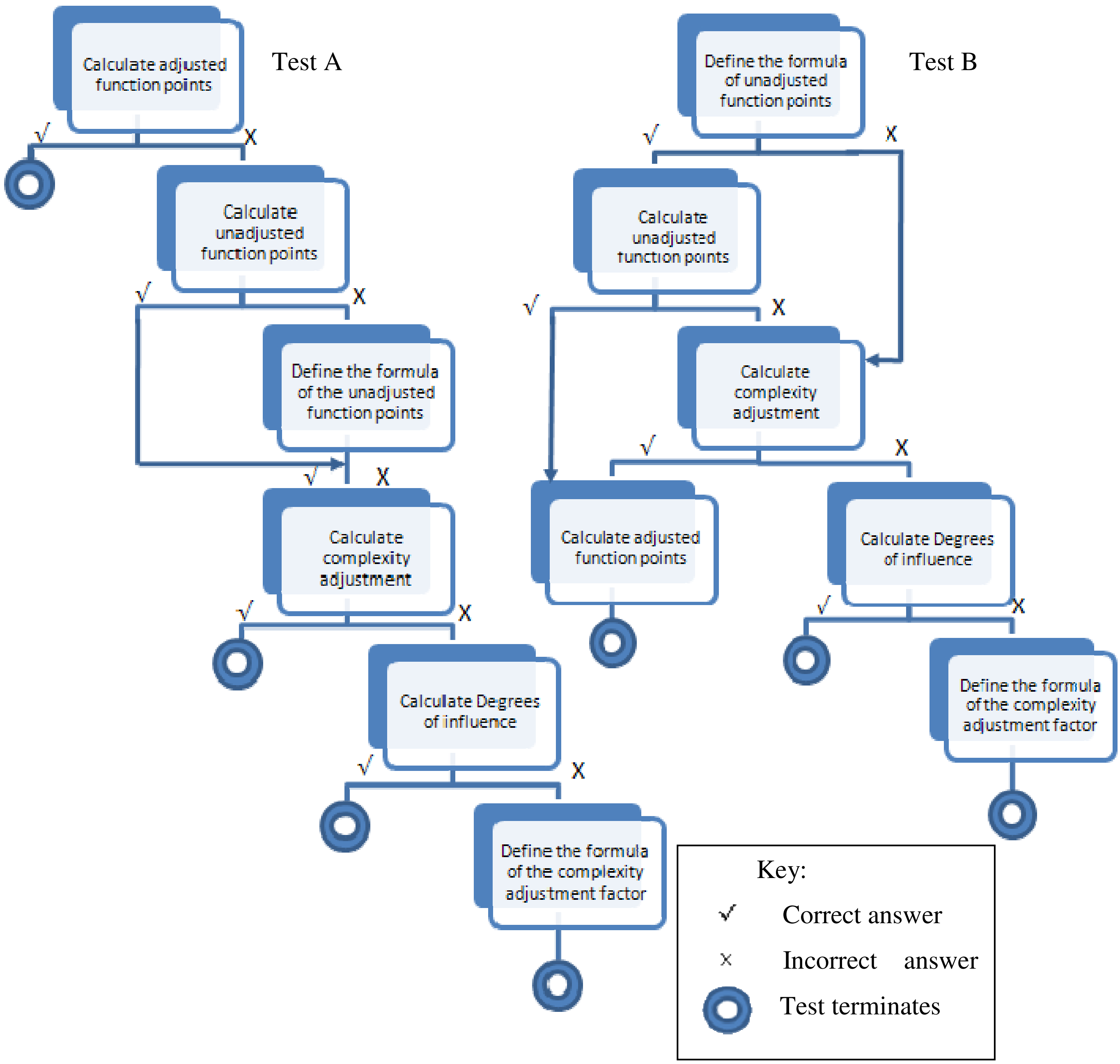
Adaptive testing aims to assess a student's knowledge by posing a number of different questions, so students are able to identify what they have already learned, to observe their personal learning progress, and to decide how to further direct their learning process.

This questionnaire asks you to evaluate four different sequences of questions in an adaptive test corresponding to the learning outcome; 'students can calculate adjusted function points'. In order to achieve this learning outcome, students would calculate the unadjusted function points and complexity adjustment.

The 'Function points' topic involves associated issues as follows:

- adjusting the function points,
- unadjusted function points,
- complexity adjustment,
- degrees of influence,
- the formula for unadjusted function points,
- estimating function points from an ER Diagram,
- calculating unadjusted function points from
    - number of attributes,
    - number of relationship lines, and
    - number of data entities.

Please answer the following questions by ranking the test which most closely reflects your opinion.



PART 1: Comparing between test A and B, which test does the most

1.1 **fairly assess the level of a student's knowledge?** Test A Test B

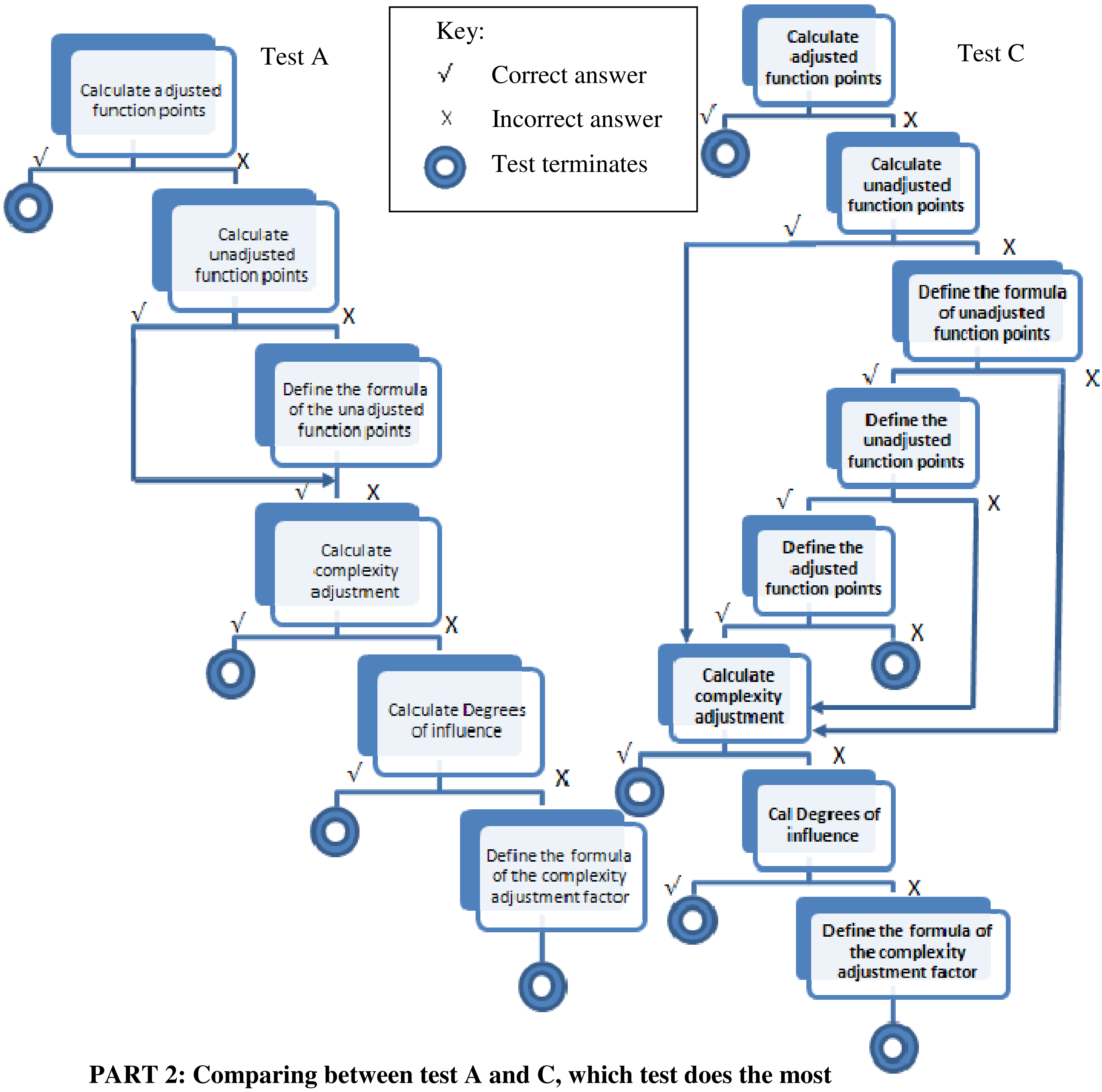
1.2 **help a student to identify what they do not know?** Test A Test B

1.3 **help a student to understand how a given learning outcome decomposes into learning outcome components?** Test A Test B

1.4 **help a student to understand how a given learning outcome decomposes into topics?** Test A Test B

1.5 **provide a complete assessment of the level of a student's knowledge for a teacher?** Test A Test B

1.6 **usefully provide for a student's own self-assessment?** Test A Test B



PART 2: Comparing between test A and C, which test does the most

1.1 **fairly assess the level of a student's knowledge?** Test A Test C

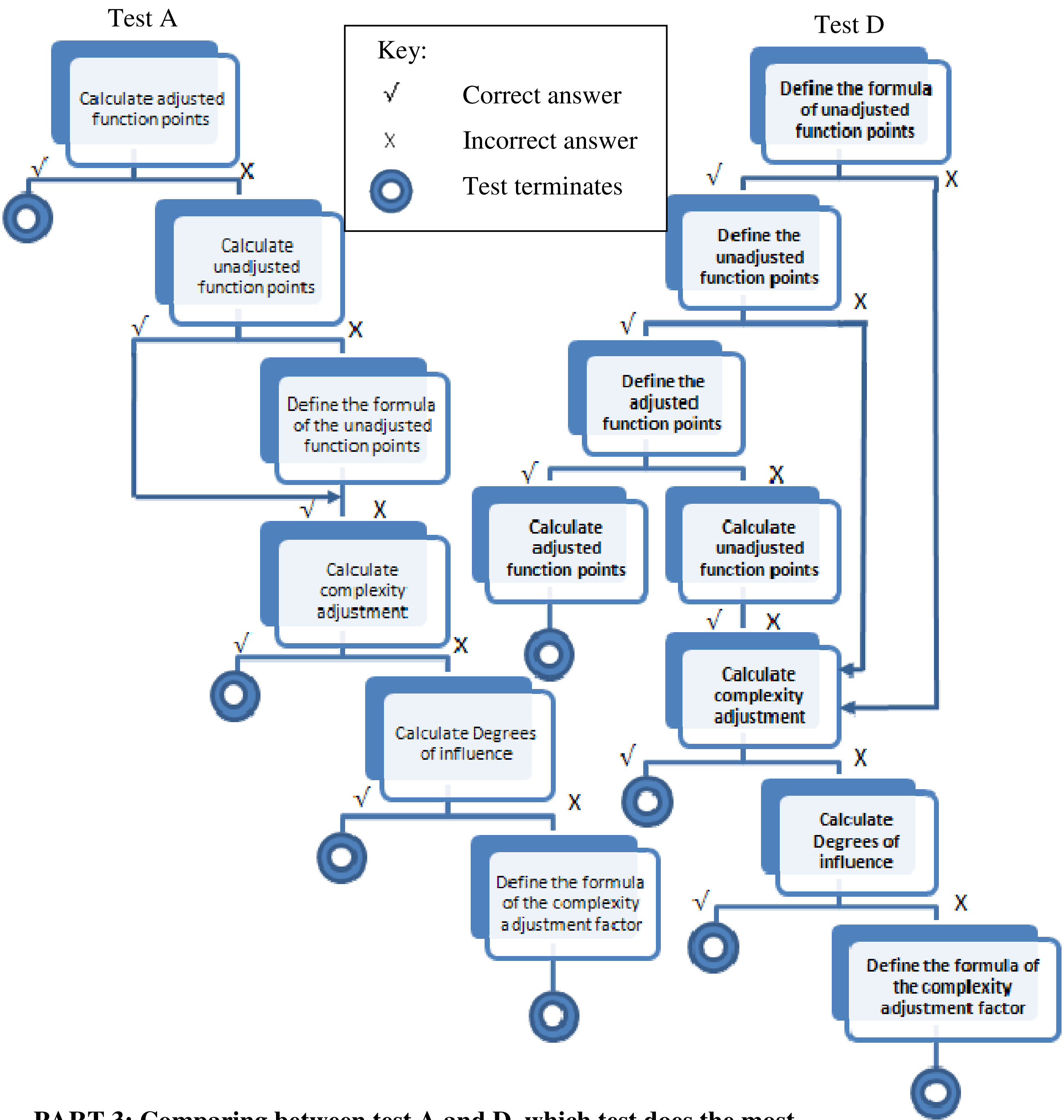
1.2 **help a student to identify what they do not know?** Test A Test C

1.3 **help a student to understand how a given learning outcome decomposes into learning outcome components?** Test A Test C

1.4 **help a student to understand how a given learning outcome decomposes into topics?** Test A Test C

1.5 **provide a complete assessment of the level of a student's knowledge for a teacher?** Test A Test C

1.6 **usefully provide for a student's own self-assessment?** Test A Test C



PART 3: Comparing between test A and D, which test does the most

1.1 **fairly assess the level of a student's knowledge?** Test A Test D

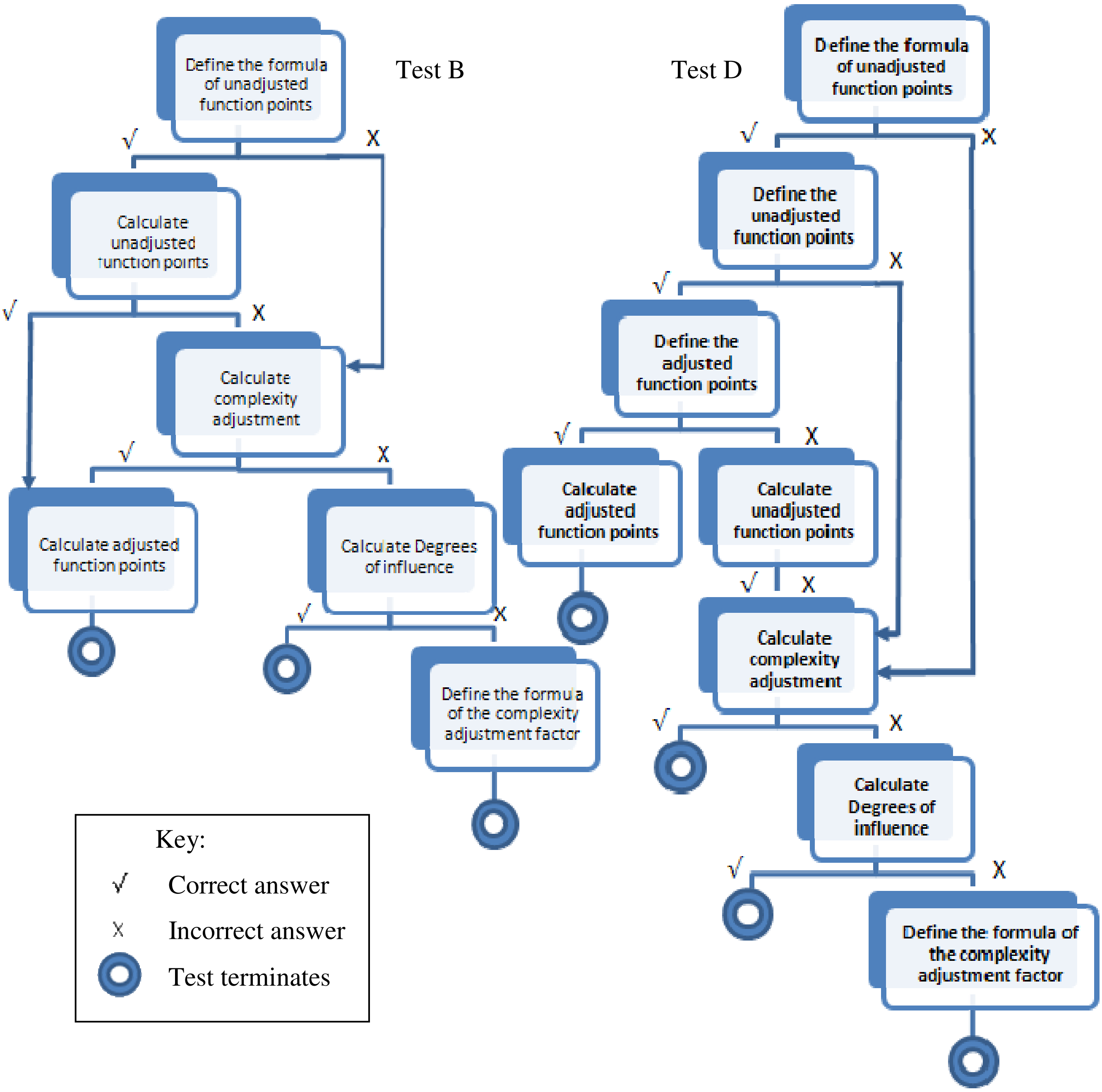
1.2 **help a student to identify what they do not know?** Test A Test D

1.3 **help a student to understand how a given learning outcome decomposes into learning outcome components?** Test A Test D

1.4 **help a student to understand how a given learning outcome decomposes into topics?** Test A Test D

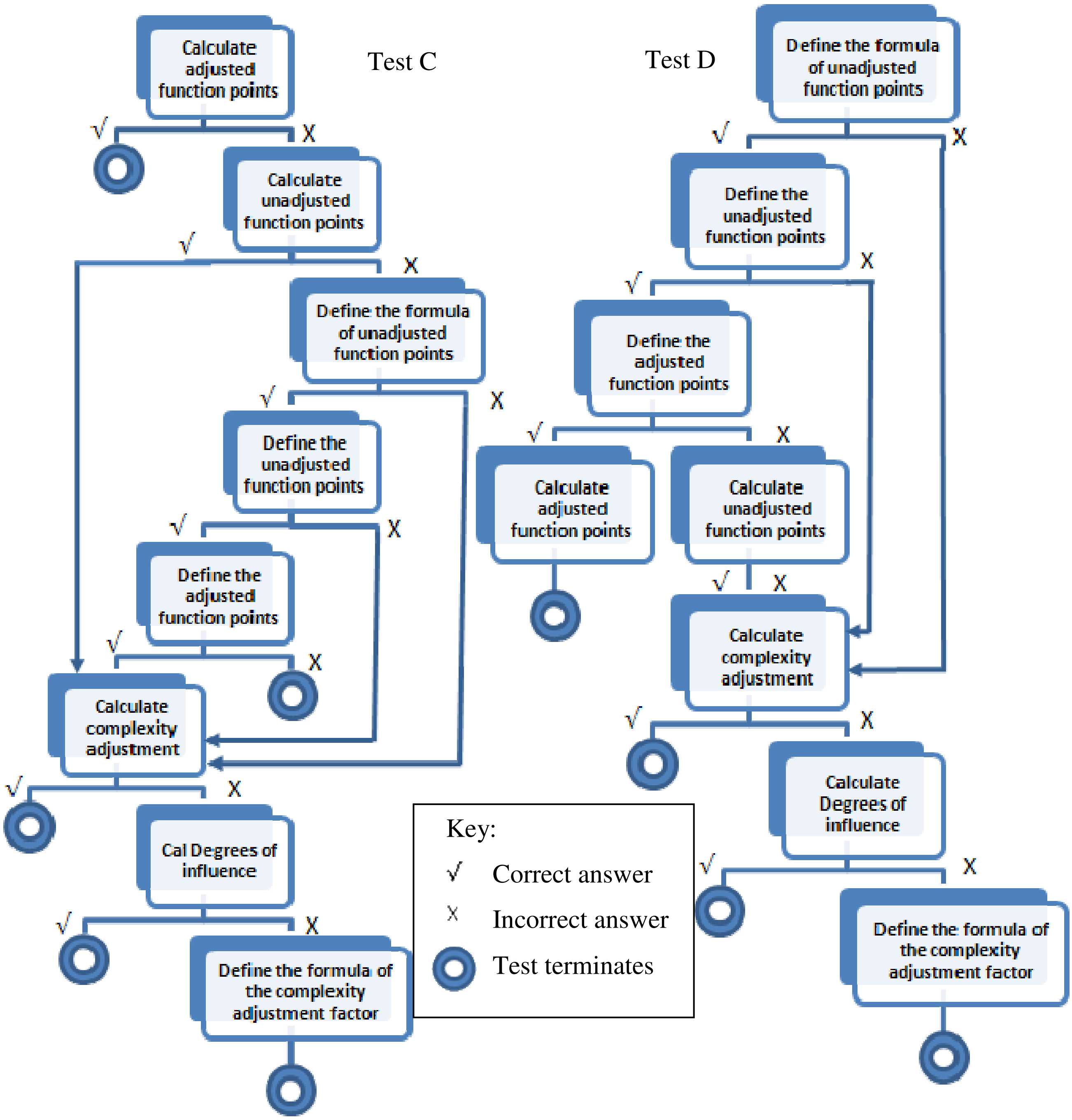
1.5 **provide a complete assessment of the level of a student's knowledge for a teacher?** Test A Test D

1.6 **usefully provide for a student's own self-assessment?** Test A Test D



PART 4: Comparing between test B and D, which test does the most

1.1	**fairly assess the level of a student's knowledge?**	Test B	Test D
1.2	**help a student to identify what they do not know?**	Test B	Test D
1.3	**help a student to understand how a given learning outcome decomposes into learning outcome components?**	Test B	Test D
1.4	**help a student to understand how a given learning outcome decomposes into topics?**	Test B	Test D
1.5	**provide a complete assessment of the level of a student's knowledge for a teacher?**	Test B	Test D
1.6	**usefully provide for a student's own self-assessment?**	Test B	Test D



PART 5: Comparing between test C and D, which test does the most

1.1 **fairly assess the level of a student's knowledge?** Test C Test D

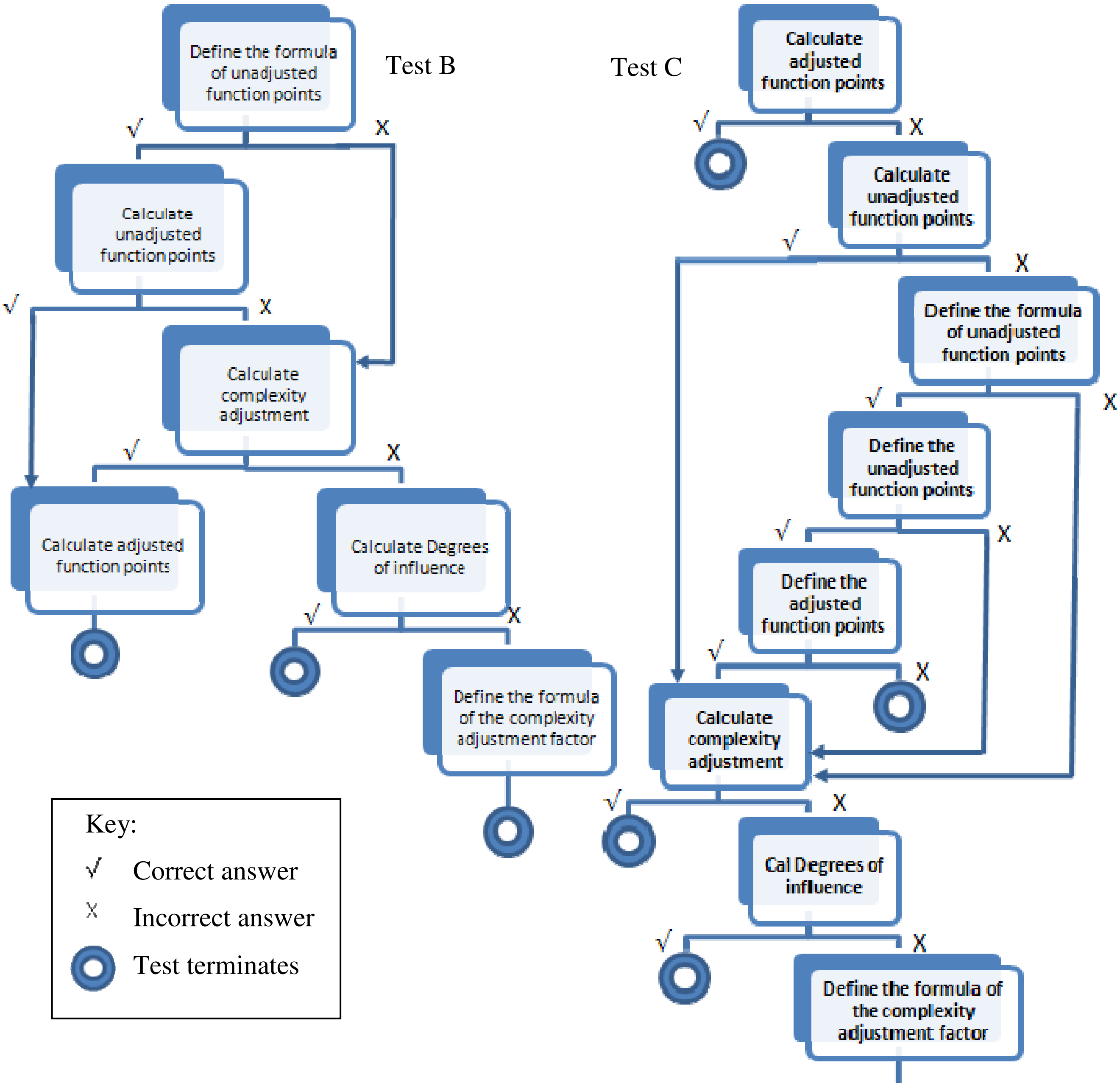
1.2 **help a student to identify what they do not know?** Test C Test D

1.3 **help a student to understand how a given learning outcome decomposes into learning outcome components?** Test C Test D

1.4 **help a student to understand how a given learning outcome decomposes into topics?** Test C Test D

1.5 **provide a complete assessment of the level of a student's knowledge for a teacher?** Test C Test D

1.6 **usefully provide for a student's own self-assessment?** Test C Test D



PART 6: Comparing between test B and C, which test does the most

1.1 **fairly assess the level of a student's knowledge?**

Test B Test C

1.2 **help a student to identify what they do not know?** Test B Test C

1.3 **help a student to understand how a given learning outcome decomposes into learning outcome components?** Test B Test C

1.4 **help a student to understand how a given learning outcome decomposes into topics?** Test B Test C

1.5 **provide a complete assessment of the level of a student's knowledge for a teacher?** Test B Test C

1.6 **usefully provide for a student's own self-assessment?** Test B Test C