Designing an Intelligent Question Design Information System (IQDIS)

Dr Elozor Shneider, The Open Polytechnic of New Zealand, New Zealand
Dr Olga Gladkikh, Tertiary Education Commission, New Zealand

Abstract

The primary objective of the research project described in this paper was to improve and simplify the assessment design process by building an intelligent information system, "IQDIS" (Intelligent Question Design Information System) for generating assessments. The key features of the system include rationalisation of assessment tasks based on both a modified cognitive taxonomy approach and a built-in mechanism for assessment quality assurance that includes an analysis of the allocation of marks. This paper discusses the main results obtained in the first project phases: determining the conceptual platform and developing a quantitative approach to quality assurance of assessments. It also describes the main functions of the new information system.

To determine the conceptual foundation, we considered a layered approach to the design of questioning strategies based on a hierarchical model for the cognitive domain, commonly known as 'Bloom's taxonomy'. We have further developed the cognitive model, introducing a third dimension to the taxonomy. This modification allows for reflection on the relative difficulty of questions pertaining to different cognitive levels, and also allows for quantitative evaluation of an assessment.

We have also developed an algorithm for evaluating the distribution of marks for assessment tasks. This algorithm enables analysis of trends in mark allocation in respect of different cognitive categories and evaluation of the impact of each category on assessment design. This algorithm has underpinned significant progress towards quantification of the quality assurance of assessment question sets.

The suggested approach enables assessment designers to perform accurate quantitative comparisons and evaluations of assessments that include such key elements as the coverage of learning objectives, addressing different levels of the cognitive process, the variety of question types, and ensuring fairness of mark allocation.

Introduction

In recent years the demands for improved quality of education and value for money have become key issues and they are now the main goals of educational institutions and professional academics. Developing high quality, robust and comprehensive assessment tools is instrumental in achieving such goals.

Manual design of assessments is a difficult and time-consuming process. For many years great effort has been put into developing information systems capable of generating assessments with different degrees of automation, e.g., Li & Sambasivan (2005), Papasalouros et al (2008), Zualkernan et al (2009). In our opinion, the greatest challenges in developing such systems are as follows:

1. Optimisation of question selection so that assessment questions target different cognitive levels within subject domains

2. Developing quality assurance process to ensure fairness and consistency of the assessments.

In our research we have attempted to create a comprehensive conceptual platform based on achievements in pedagogical theory and psychology of learning that provides a solid foundation for developing assessments and building a new generation of automatic assessment composition systems.

This research project aims to improve and simplify the assessment design process by building an intelligent information system 'IQDIS' (*Intelligent Question Design Information System*) for generating assessments. The key features of this system include rationalisation of assessment tasks based on a modified cognitive taxonomy and a built-in mechanism for assessment quality assurance that analyses the allocation of marks. This paper presents an overview of the results obtained in the first phases of the project: determining the conceptual platform and developing a quantitative approach to quality assurance of assessments. It also describes the main functions of the new information system.

Determining conceptual platform

We applied the following criteria to judge the appropriateness of a conceptual platform:

- It should address a range of students' abilities to express their thoughts
- It should be scalable both horizontally (i.e. within a topic coverage) and vertically (i.e. across all levels of mastery).

These criteria suggest a layered model type as a possible solution. In Shneider & Gladkikh (2006a, 2006b), we considered a layered approach as a first step towards the design of questioning strategies, based on Anderson's interpretation of the hierarchical model for the cognitive domains (Anderson & Krathwohl, 2001). The original model is commonly known to educators as 'Bloom's taxonomy' or 'Bloom's model of critical thinking' (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956). This approach represents a hierarchy of cognitive layers as follows:

Recall   Comprehension   Application   Analysis   Evaluation   Synthesis

Figure 1: One-dimensional representation of the taxonomy model. This is a slightly modified representation of Bloom's taxonomy reflected in Shneider & Gladkikh (2006b).

Krathwohl (2002) and Anderson (2005) suggested the addition of a second (Knowledge) dimension to the cognitive process model. The Knowledge Dimension comprises four subcategories that are designed to reflect the following ideas:

A. Factual Knowledge – the basic elements that students must know in order to be acquainted with a discipline or to solve problems within it.

    Aa. Knowledge of terminology
    Ab. Knowledge of specific details and elements.

B. Conceptual Knowledge – the interrelationships among the basic elements within a larger structure that enable them to function together.

    Ba. Knowledge of classifications and categories
    Bb. Knowledge of principles and generalizations
    Bc. Knowledge of theories, models and structures.

C. Procedural Knowledge – how to do something; methods of inquiry, and criteria for using skills, algorithms, techniques and methods.

Ca. Knowledge of subject-specific skills and algorithms

Cb. Knowledge of subject-specific techniques and methods

Cc. Knowledge of criteria for determining when to use particular procedures.

D. Metacognitive Knowledge – knowledge of cognition in general, as well as awareness and knowledge of one's own cognition.

Da. Strategic knowledge

Db. Knowledge about cognitive tasks, including appropriate contextual and conditional knowledge

Dc. Self-knowledge.

The two-dimensional (2D) model of the cognitive process implements a synthetic approach to knowledge delivery and assessment and allows for finer differentiation within a cognitive level. However, this model does not account for the degree of similarity of question types within the 2D cognitive cell. It also does not reflect the relative difficulty of questions that belong to different cognitive levels and does not allow for quantitative evaluation of an assessment.

To resolve these issues we suggested the addition of a third dimension to the taxonomy (Shneider & Gladkikh, 2007). For clarity we refer to the new version of the Modified Taxonomy as 3D-MT.

The new dimension in 3D-MT represents the level of effort required to answer a given question. We refer to this new parameter as the level of difficulty. A ***level of difficulty*** is a composite value that embodies two components: the ***rank of difficulty (difRank)*** of a question within a given cognitive domain and a ***cognitive coefficient (cognCoef(N))*** that reflects the rising complexity of questions at higher cognitive levels.

The first component of the new dimension (***difRank***) estimates the level of effort required to answer the question within a given cognitive domain (e.g. difficulty scale 1 – 5). It refers to the required summative level of knowledge estimated on the basis of a knowledge dimension parameter (A...D in Krathwohl's (2002) description). To estimate ***difRank*** as a first approximation, an arbitrary categorical approach to knowledge-based difficulty could be used. However, in the future, a more accurate quantitative approach to evaluation of questions' difficulty based on Item Response Theory (Baker & Kim, 2004) will be applied to the bank of questions.

The second component, the ***cognitive coefficient***, has been modelled (Shneider & Gladkikh, 2007) using the following assumptions:

1. The cognitive coefficient rises from Recall to Synthesise
2. It shows smooth growth which is slower at higher cognitive levels
3. It ranges between 1 and 3
4. The ratio of cognitive coefficients for any two neighbouring cognitive levels is less than 2.

Based on the above assumptions, the following empirical formula for calculating the cognitive coefficient *cognCoef(N)* has been introduced:

$$cognCoef(N) = 3.3 * \log(1 + N) \qquad (1)$$

where *N* is the cognitive level (Recall to Synthesise correspond to the values of N 1 to 6); ***3.3*** – an empirical coefficient adjusted so that the function meets the requirements. Figure 2 below illustrates the behaviour of the obtained empirical function *cognCoef(N)*:

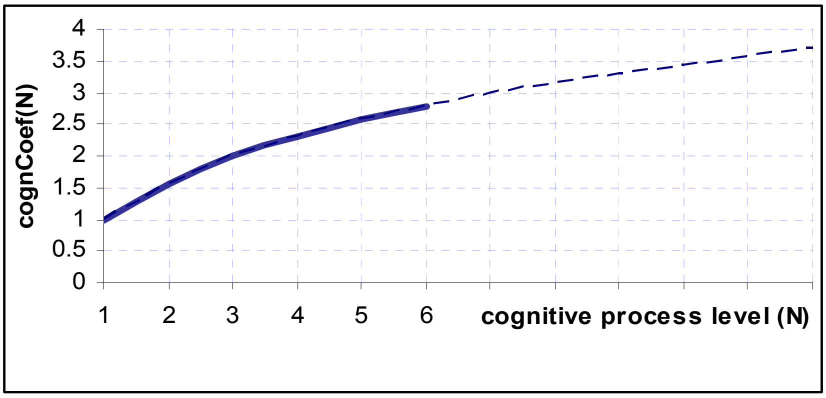

Figure 2: The cognitive coefficient vs. cognitive level.

According to formula (1), the value of *cognCoef* increases from 1 (Recall) to 2.8 (Synthesise).

For each question in an assessment set, a level of difficulty can be calculated as a Cartesian product of the difficulty rank and the cognitive coefficient at the level of the cognitive dimension:

$$levelDifficulty = difRank(q)*cognCoef(N) \qquad (2)$$

where q represents the question identifier (see below) and N represents the intended cognitive level of the question. To describe a difficulty level for the entire assessment set, a summative value of levels of difficulty for each question should be used. We propose the following:

$$OverallDifficultyRating = (\sum difRank(q)*cognCoef(N))\,/n \qquad (3)$$

where n is the total number of the assessment questions.

The following additional parameters were also used to describe a question within the 3D-MT taxonomy:

Question identifier (*q*) - an alphanumeric ‘address’ of a question within an assessment. For example, **q = 31a** refers to question 1a from part 3 of the test.

Question type (*type index*) - a numerical value identifying a definitive combination of keywords suggested for each cognitive domain. An example of the type index assignment is shown in Table 1.

Cognitive process	**Problem type**	**Specific key word**	**Knowledge Dimension** (Krathwohl, 2002)	**Type index**
Apply	***Write*** code to declare an array with given specifications	Write	Bc	1

	Amend the code so that the elements of an array are printed out in a required format	Change	Cb	2
	Apply a sorting procedure to a given array containing N random numbers to arrange the elements in ascending order	Apply	Ca	3
	Write a procedure to place the numbers from a given array into the other arrays	Write	Cc	4
	Create an application that displays an average, median, maximum and minimum exam scores based on the user input	Create	Cb	5

Table 1: An example of question type index assigned at the Apply level.

Figure 3 summarises the general concept for the 3D-MT model and Figure 4 presents an aggregate table for the 3D-MT model.

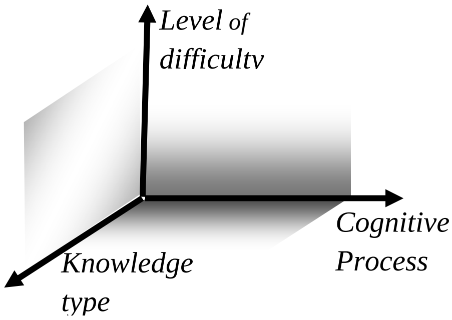

Figure 3: Three - dimensional presentation for the 3D-MT model.

The following example (cell A4 in Figure 4) contains one entry, 31a(1,2,3), where 31a is a question identifier, the question type is 1, the rank of difficulty is 2, and the allocated mark for this question is 3.

	Recall	**Comprehend**	**Apply**	**Analyse**	**Evaluate**	**Synthesise**
N:	1	2	3	4	5	6
Cognitive coefficient ***cognCoef:***	1	1.6	2	2.3	2.6	2.8
The Knowledge Dimension						
A. Factual Knowledge				***31a(1,2,3)***		
B. Conceptual Knowledge						
C. Procedural Knowledge						
D. Metacognitive Knowledge						

Figure 4: Cognitive thinking 3D-MT model.

Assessment quality assurance
To evaluate fairness of marks allocation for an assessment, we suggested the following algorithm (Shneider & Gladkikh, 2007):

1. Analyse the distribution of difficulties for questions at each cognitive level in a given assessment set

- Group the questions by cognitive level ***(N)***. The total number of such groups will be six or less (the number of cognitive levels).
- Calculate the sum of the Cartesian products of the difficulty ranks and the cognitive coefficient for each group. The obtained value will show the summative level of difficulty for all questions at a particular cognitive level:

$$\sum_{q} (difRank(q)*cognCoef(N))$$

- Add summative difficulties for each cognitive level **N** calculated at the previous step to find a summative difficulty for the whole assessment. Divide the obtained number by the total value of allocated marks for the assessment. Using the obtained ratio, scale the summative difficulty at each cognitive level of the assessment to enable comparison of difficulty distribution with the marks distribution. This step is optional; non-scaled values could be used for the same purpose.
- Plot the resulting distribution of question difficulties ∑***(difRank(q)*cognCoef(N))*** vs ***N*** . An example of this distribution is represented by curve 1 (see Figure 5).

2. Analyse the actual mark distribution according to the marking schedule for the same question group

- Use the same question grouping (by cognitive level) as in p.1. Sum the marks allocated for each group of questions.
- Plot the distribution of mark allocations ***levAllocateMarks*** vs ***N***. An example of such distribution is depicted in curve 2 (see Figure 5).
- Compare the gradients of graphs 1 and 2. If the main trends (rises and falls) for both graphs correspond (i.e. the product of the signs of the gradients is positive), the trend in the actual mark distribution is correct.

If (Sign(grad(∑***(difRank(q)*cognCoef)***))* Sign(grad***(levAllocateMarks)***)) > 0 The trend in the mark allocation is correct ***Else*** The trend in the mark allocation is incorrect

- Compare the deviation of values of ***levAllocateMarks*** from ∑***(difRank(q)*cognCoef)*** shown in graphs 1 and 2 for each ***N.*** Ideally, for a non-scaled distribution of questions' difficulties the deviation should be a constant value for all ***N***, with little discrepancy. For a scaled distribution of estimates, the ideal deviation should be close to zero with non-significant discrepancy. The presence of outliers would suggest that some adjustment of the corresponding marks is needed:

If (***levAllocateMarks(N) - ∑(difRank(q)*cognCoef)***) **<** ***setLimit*** Mark allocation is correct ***Else*** Mark allocation needs adjustment

Figure 5 shows the evaluation of allocated marks for a sample examination (for a course in computer programming). The assessment included questions that addressed all six cognitive levels, and therefore each graph covers six points along the horizontal axis. The vertical axis represents the marks (lower curve, curve 2) and calculated difficulties (upper curve, curve 1) associated with the groups of questions at each cognitive level. Although graph 2 shows that the assessment set was intended to focus on three levels (Recall, Analyse and Evaluate) in respect of the assessed subject (***N* = *1, 4*** and **6**), the distribution of questions difficulties (graph 1) clearly shows weighting towards the Analyse level (***N* = *4***), where the allocated marks did not reflect the level of difficulty for this group of questions properly. The presence of an outlier at ***N* = *4*** means that either the marks for the Analyse group of questions need to be adjusted and scaled up or else that the level of difficulty of the Analyse questions should be adjusted to match those at the Recall and Evaluate levels.

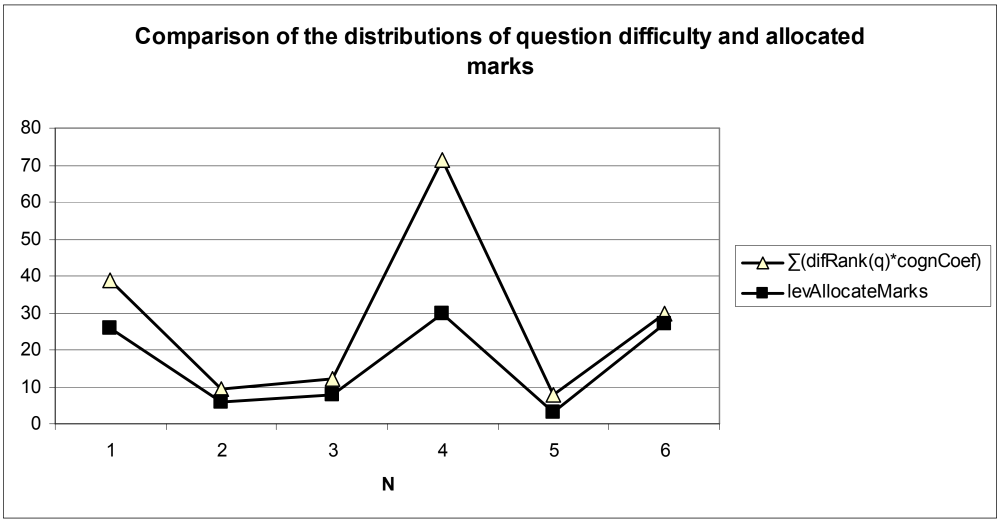

Figure 5: The distribution of question difficulties (upper curve 1) and actual mark allocation (lower curve 2) for the sample examination.

Conceptual design for IQDIS

The previous sections outlined a conceptual platform for the intended information system for generating assessments 'IQDIS' (Intelligent Question Design Information System) and identified the system's main capabilities. In order to implement those capabilities, IQDIS should embody the following basic functions:

- Manage the question bank database (add new questions to a question bank, edit questions, delete questions)
- Create new assessments (sets of questions) based on given criteria. Criteria for selection of questions for a new assessment could include such elements as the subject of assessment, topic of assessment, question form (e.g. multi-choice, essay, problem solving etc), preset level of difficulty (e.g. easy, moderate, difficult, mixed level), cognitive dimension level, knowledge type, similarity to an existing assessment (if stored), and the number of questions of each type
- Edit/modify an assessment
- Display the marking schedule and model answers
- Evaluate existing assessments against a model distribution of marks and against other assessment. Evaluation criteria may include:

a) Coverage of learning objectives
b) Addressing different levels of cognitive process
c) Variety of question types
d) Fairness of mark allocation. Criteria for fairness of mark allocation could include such elements as whether a distribution of marks for overall assessment (and/or marks for individual questions) falls within the set limit of tolerance. A set limit of tolerance in mark allocation (actual vs. modelled) should be specified (for example, 10%)
e) Overall level of difficulty. This value is calculated using the cognitive coefficient and a difficulty rank of questions for a given topic within the cognitive domain using formula (3).

The identification of such functions enabled us to perform the conceptual level design for IQDIS. A conceptual data model should encompass all data requirements of the system and support all system functions. It should also be used to determine the system's User Interface. Figure 6 shows a conceptual data model for IQDIS that represents the relevant elements of the problem domain and the relationships that must be captured.

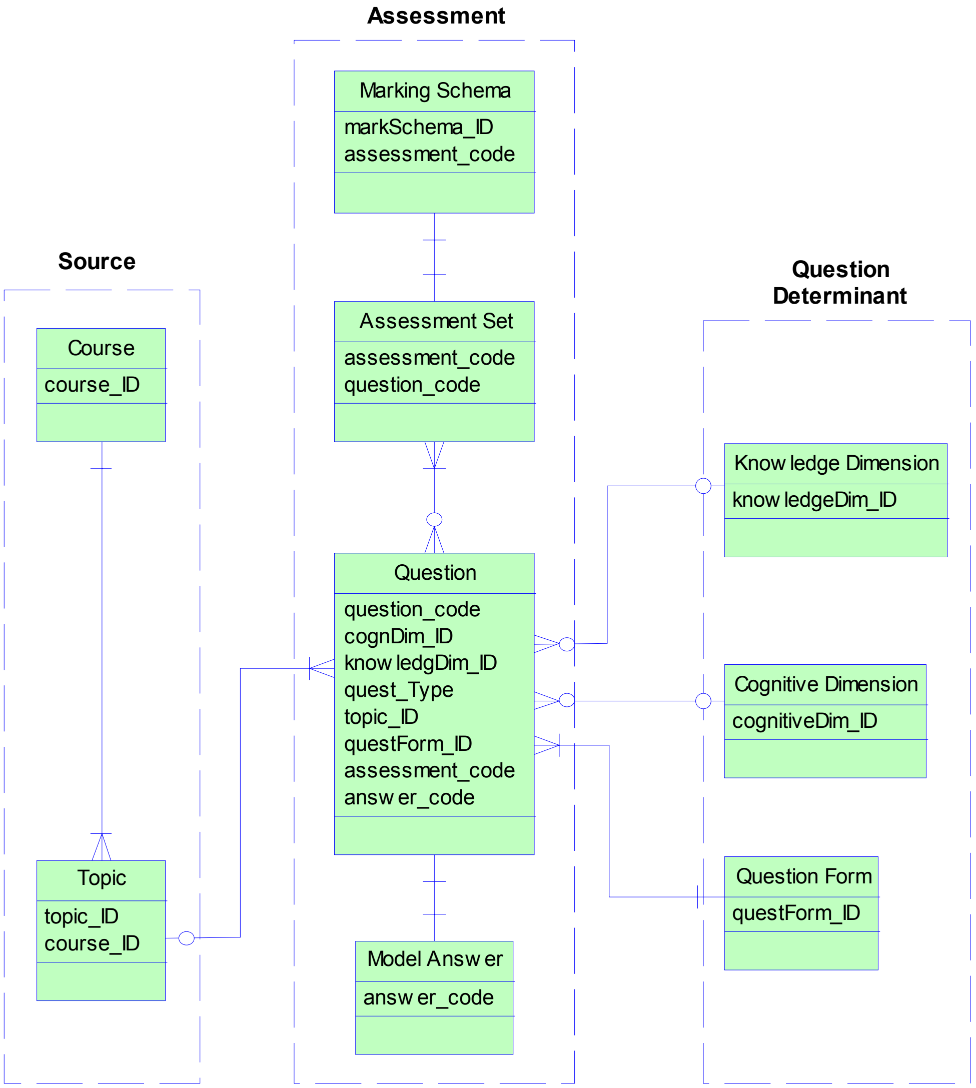

Figure 6: Conceptual entity-relationship diagram for IQDIS.

The conceptual level schema for data storage in IQDIS includes three basic domains – Source, Assessment and Question Determinant. The domain Source currently includes entities Course and Topic. They represent the knowledge base for each question in the assessment set and are referenced by unique identifiers. Apart from the unique identifier, the attributes of Course may include the course name and descriptor. In addition it may also include relevant organisational codes such as a unique code used by government ministries for the subject (if applicable), the number of hours required and (optional) a coefficient showing the contribution of the course to a particular qualification (if known) etc. Most of the government accredited courses have well-defined learning outcomes that may serve as attributes of the Course entity as well. The entity Topic is also described by its name, unique code and a descriptor. Each instance of a Course entity contains a list of topics, and thus the cardinality of the relationship between these two entities is set as one to many.

The domain Question Determinant consists of entities that describe the form and cognitive context of a question, together with the type(s) of knowledge addressed by the question. Thus, it includes the entities Cognitive Dimension, Knowledge Dimension and Question Form. The entity Cognitive Dimension includes such attributes as cognitive level (as in the 3D-MT model), cognitive level description, key-words typical for each level and a cognitive coefficient obtained using formula (1). At this stage, the entity Knowledge Dimension is described by a unique reference to a particular knowledge dimension used across the database (the symbols from Krathwohl, 2002 were used) and knowledge dimension description. The entity Question Form includes such attributes as form code, form type (e.g. multi-choice, essay, problem work out etc) and the expected form of answer. The range of values for the latter may include the number of options (for optional choice questions), the number of correct options (for check-box type questions), component formula for essay (e.g. definition only, definition and explanation, definition and explanation and examples etc), answer or answer and workout etc. All three entities are related directly to the entity Question from the Assessment domain.

The domain Assessment includes entities representing assessment itself and its components – Assessment Set, Marking Schema, Questions and Model Answers. The entity Question represents the hub of the conceptual schema, being related to six other entities belonging to three conceptual domains. Thus, it contains several external identifiers to be used as foreign keys. The other attributes are a unique code, a question descriptor and the knowledge depth-based rank of difficulty assigned to each question within a cognitive level. This entity also includes a flag field indicating that this question has already been processed and included in an assessment.

The entity Model Answer includes an answer identifier in the database, the question code to which the answer is related, a generic answer to the question, a measure of the level of ambiguity that shows the possibility of different readings of the same question, alternative algorithms for problem solution and relevant answers. The functionality of IQDIS will enable the generation of concrete (instead of generic) answers that would include either interaction with the user or generating a random choice from available range of input data for a question.

The entity Assessment Set represents an assessment as a single object with its unique identifier, number of parts, number of questions included in each part, assessment category (e.g. diagnostic, formative, summative), calculated overall level of difficulty and assessment type (interactive, write-on) etc. It will also include an optional field for the weighting of this assessment within a total course assessment. The entity Marking Schema will describe the mark allocation for a particular assessment. Thus, it will include the assessment code for which it was created, the question identifier, the number of marks allocated for each question and several calculated fields such as marks per assessment section and the total for an assessment.

Conclusion

Targeting of the assessment questions to different levels of cognitive thinking can improve assessment design significantly and enhance the efficiency of teaching and learning. This paper has summarised the main results achieved to date in the IQDIS project: determining the conceptual platform; developing a quantitative approach to quality assurance of assessments; developing a conceptual design model for the system. Adding a third dimension - the level of difficulty - to the cognitive taxonomy allows for quantitative evaluation and comparison of assessments, as well as the analysis of trends in mark allocation in relation to various cognitive categories. The suggested approach serves as a step towards the quantification of the quality assurance of educational assessments.

References

Anderson, L. (2005). Objectives, evaluation, and the improvement of education. *Studies in Educational Evaluation*, *31*, 102-113.

Anderson, L.W. & Krathwohl, D. (Eds.). (2001). *Taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Baker, F.B., & Kim, S.-H. (2004). *Item Response Theory (2nd Ed).*

Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive Domain*. New York: Longmans Green.

Krathwohl, D. (2002). A revision of Bloom's Taxonomy: An overview. *Theory Into Practice*, *41*(4), 212-218.

Li, T., & Sambasivan, S. (2005). *Automatically Generating Questions in Multiple Variables for Intelligent Tutoring*. Issues in Informing Science and Information Technology. Retrieved July 23, 2009. from http://informingscience.org/proceedings/InSITE2005/I39f57Li.pdf

Papasalouros, A., Kanaris, K., & Kotis, K. (2008). *Automatic Generation of Multiple Choice Questions from Domain Ontologies.* Retrieved July 23, 2009 from http://www.icsd.aegean.gr/kotis/publications/IADIS2008.PDF

Shneider, E. & Gladkikh, O. (2006a). Designing questioning strategies for information technology courses. In S. Mann & N. Bridgeman (Eds.), *Proceedings of the 19th Annual Conference of the National Advisory Committee on Computing Qualifications* (pp. 243-248). Wellington, New Zealand: NACCQ.

Shneider, E. & Gladkikh, O. (2006b). Designing questioning strategies: applying Bloom's taxonomy to online computer programming courses. In J.R. Parker (Ed.), *Proceedings of the Second IASTED International Conference on Education and Technology* (pp. 194-199). Calgary, AB, Canada: IASTED.

Shneider, E. & Gladkikh, O. (2007). Developing quality assurance metrics for assessment design. . *New Zealand Journal of Applied Computing and Information Technology*, *11*(1), 53-58.

Zualkernan, I. A., El-Naaj, S. A., Papadopoulos, M., Al-Amoudi, B. K., & Matthews, C. E. (2009). Automatic Generation of Justin-time Online Assessments from Software Design Models. *Educational Technology & Society*, *12* (1), 173–192.