T. Hirashima et al. (Eds.) (2011). Proceedings of the 19th International Conference on Computers in Education. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education

Cognitive Assessment Applying with Item Response Theory

Onjira SITTHISAK[a*], Tasanawan SOONKLANG[b] & Lester GILBERT[c]
[a] *Computer and Information Technology Group, Faculty of Science, Thaksin University, Thailand*
[b] *Department of Computing, Faculty of Science, Silpakorn University, Thailand*
[c] *Learning Societies Lab, School of Electronics and Computer Science, University of Southampton, United Kingdom*
*onjira@tsu.ac.th, tasanawa@cp.su.ac.th and lg3@ecs.soton.ac.uk

Abstract: Using item response theory (IRT) in computer adaptive test (CAT) is critical to assess learners' readiness for further learning. This theory has assumptions concerning the mathematical relationship between learner's abilities and item responses. A numerical value from the theory has potential to decide who the best learner is. The IRT score only provides the learner's result and lack of the assessment of learner's ability level in cognitive skills. This paper proposes an approach of using Bloom's cognitive skills for evaluating learned ability level by considering achieved the level of difficulty and the IRT score. Our starting point in the adaptive test is at a question of learners' knowledge level difficulty that we got from their pre-test scores instead of starting with a question of a medium difficulty. This approach provides the effectiveness of CAT to test cognitive skills and offers some theoretical considerations on linking learning outcomes and assessments.

Keywords: item response theory, computer adaptive test, cognitive skill

Introduction

There are a number of computer adaptive test 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 testing changes their behaviour and structure depending on the learner's responses and detected abilities.
The key idea of an adaptive test 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 [1,2]. Adaptive testing aims to assess a learner's ability by posing a minimum number of questions in order to decrease test length, which is one of the main goals in adaptive testing [3]. Another main goal includes offering personalized support according to the needs and ability of each learner [4]. 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 [1]. Therefore, adaptive questioning is an efficient and effective mean of knowledge-based testing.
Many adaptive testing systems have been developed such as A Web-based English CAT prototype system [5], IDEAL [6], and SIETTE [7]. Item response theory (IRT) was used in the implementation of these systems because the proficiency estimate is independent of the particular set of questions selected for the assessment. Each learner gets a different set of questions with different difficulty levels while taking the adaptive testing [8].

These systems are focused on using IRT to estimate the numerical value of learner's ability level, in order to determine the next item to be posed, and to decide when to finish the test, rather than to assess learners' readiness for further learning. One of the major challenges facing the use of IRT is establishing standards for usability and interpretability issues of the IRT value [9]. In IRT, ability is measured by a scale point. When applied this theory to measure cognitive skills expected to be tested in each learning outcome such as Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation, the theory has some limitations [10]. In this paper, we introduce IRT and explore cognitive assessment. Finally, the proposed cognitive adaptive testing system applying with IRT is discussed and the main contributions are summarized.

1. A Proposed Cognitive Adaptive Testing

We propose to build an adaptive testing system by applying with IRT where teachers can create, manage, and control their test for each learner. The system would be a cognitively-based formative assessment that will identify more detail about the underlying concepts and skills needed to solve problems in each reporting category. Each test, the system will learn more about each learner and provide an increasingly accurate prediction of how well each learner will do on their test.

One of the objectives is to add a new factor to the IRT model which is a pre-test score. We use this factor for initial learner ability in order to tailor to learners' ability level instead of starting a question with a medium difficulty. A pre-test will measure a learner's knowledge level before starting a real adaptive testing.

Table 1 Guidelines for expert calibration

Difficulty(b)	Cognitive skill	Skill being assessed
-3 <= b<= -2	Remembering	Ability to recall taught material
-2 <= b<= -1	Understanding	Ability to interpret and/or translate taught
-1 <= b<= 0	Applying	Ability to apply taught material to novel situations
0 <= b<= +1	Evaluating	Ability to make judgments based on criteria & standards
+1 <= b<= +3	Creating	Ability to reorganize elements into a new pattern

A proposed system prototype comprises a graphical user interface, an adaptive algorithm based on the 3-PL model from IRT and item bank questions. The item bank of questions is employed to store information about question stem, distractors, key answers, topic area and values for the parameters required by the 3-PL model [11,12]. One of the central elements of the 3-PL model is the level of item difficulty (b_i) being defined by subject domain experts. The expert calibration is based on Bloom's taxonomy of cognitive skills [13] illustrated in Table 1. Our questions break down individual test items into component concepts and skills in order to track what knowledge and skill limitations of individual learners.

A proposed adaptive testing system consists of a number of modules: test editor, IRT editor, IRT calculator, pre-test generator and adaptive test generator (illustrated in **Error! Reference source not found.**). The test editor is responsible for teacher to create questions based on cognitive skill theory, and storing the questions in a test bank. In order to estimate IRT, the IRT parameters such as difficulty, discrimination, and pseudo-guessing have to be filled in and calculated by teacher, and storing in a IRT repository. Before learners start to do their adaptive test, they will start to evaluate their knowledge by using a pre-test. Our starting point in the adaptive test is at their knowledge level that we got from their pre-test scores. The adaptive test generator module retrieves a question from the test bank based on their level. If learners' response is correct, the module will recalculate their ability and show more difficulty questions to follow till the level of learners' ability becomes stable. Conversely, an incorrect response will trigger a less difficult question to be administered next.

2. Conclusion

Many adaptive testing systems have been developed to assess learner's ability level. However, the proficiency of learner from those systems was not linked to learning outcomes in terms of cognitive skills. This paper proposed an adaptive testing system based on IRT and cognitive skills. The 3-PL model and Bloom's taxonomy are used in the implementation of our system. A preliminary result from pre-test is incorporated in IRT for the estimation of the initial learner ability in order to make IRT more realistic and applicable. The dichotomous item type is focused only.

The proposed system provides the effectiveness of CAT to improve an assessment connecting with learning outcomes. In this study, Bloom's cognitive skills are used to create a list of all questions that are possible at various levels. These questions are used to test understanding and in some cases determine the degree to which learners had actually acquired the desired knowledge. Our system guides question development through the six levels of cognitive ability to stimulate critical thinking. This makes it possible to guide teachers in developing questions for learners and provide authoring templates to speed the creation of new questions for assessment.

This system also has the great advantage of providing individuals with a more detailed identification of learners' performance. The system 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 learning outcome. It is suggested that information about learning outcomes should form the basis of pedagogically-informed metadata which would be relevant to any description of content or process in learning and teaching situation.

References

[1] Askins, D. (2004). Adaptive assessment. *Toolkit - The Nuts and Bolts e-Newsletter, 3*.

[2] Way, W. D. (2005). *Practical questions in introducing computerized adaptive testing for k-12 assessments*.

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

[4] Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction, 6*(2-3), 87-129.

[5] Yu-Huo, T., Yu-Lung, W., and Hsin-Yi, C. (2008). A practical computer adaptive testing model for small-scale scenarios. *Educational Technology & Society, 11*(3), 259-274.

[6] Wang, F. H. (2006). *Application of componential irt model for diagnostic test in a standard conformant elearning system.* Paper presented at the the Sixth International Conference on Advanced Learning Technologies, Kerkrade, The Netherlands.

[7] 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, 1*(33).

[8] Weiss, D. J. (1983). *New horizons in testing: Latent trait test theory and computerized adaptive testing*.New York: NY: Academic Press.

[9] Gibbons, R. D., Immekus, J. C., and Bock, R. D. (2007). The added value of multidimensional irt models. from http://www.uic.edu/labs/biostat/articles/NCI_Didactic_Workbook.pdf

[10] Junker, B.W. and K. Sijtsma, *Nonparametric Item Response Theory in Action.* Applied Psychological Measurement, (2001). **25**(3): p. 211 - 220.

[11] Lord, F. M. *Applications of item response theory to practical testing problems*. (1980), Hillsdale, NJ: Lawrence Erlbaum.

[12] Wainer, H., Dorans, N., Eignor, R., Flaugher, B., Green, B., Mislevy, R., et al. *Computerized adaptive testing: A primer*. (2000), Hillsdale, NJ: Lawrence Erlbaum Associates.

[13] Bloom, B.S. and D.R. Krathwohl, *Taxonomy of educational objectives: The classification of educational goals by a committee of college and university examiners*, ed. H.I.C. Domain. (1956), New York: Longman.