

# Measuring Procedural Knowledge in Problem Solving Environments with Item Response Theory

Manuel Hernando, Eduardo Guzmán and Ricardo Conejo

E.T.S. Informática. Universidad de Málaga,  
Bulevar Louis Pasteur, 25. Campus de Teatinos.  
29071, Málaga, Spain.  
{mhernando, guzman, conejo}@lcc.uma.es

**Abstract.** In this paper, a new data-driven model to measure procedural knowledge is described. The model is based on Item Response Theory. The main idea behind this new model is to establish an analogy between the testing and the problem solving environment. For this purpose, we model each problem (or exercise) solution path as a directed graph where nodes are states of the problem and edges, transitions between states (i.e. the actions accomplished by the student). We can match this model with testing by seeing each node as a question and each edge as choices within the questions.

**Keywords:** Problem Solving Environments, Student Modeling, Procedural Knowledge Estimate, Item Response Theory.

## 1 Introduction

The way learners acquire knowledge has changed in the last few years with the arrival of new technologies and computer-aided systems. Individualized learning, which is more effective than traditional methods [1], can be achieved through a combination of traditional methods and computer-aided ones. In order to offer students the best strategy to acquire some concepts or skills, we need to maintain a student model that represents his/her knowledge. Using that information an Intelligent Tutor System could guide students to reach a certain goal. Updating and maintaining this model is a difficult issue in the field. Students' models have to be updated as students interact with the system, so it is necessary to infer the student's model through his/her actions.

There are different strategies for representing student models in the AIED literature; most of them assume procedural-declarative distinction. Declarative knowledge refers to the knowledge of relevant principles and concepts of a certain subject that can be applied in new tasks [2]. Procedural knowledge is the acquisition of skills related to step-by-step actions in solving problem context [3]. Declarative knowledge is usually assessed with testing systems, while procedural knowledge is mainly assessed through problem solving environments.

In testing systems, the Item Response Theory (IRT) [4] constitutes probably the most successful and well founded of all the strategies. Otherwise, most of procedural assessment strategies are based on Bayesian networks. These networks can be, in the

worst case scenario, NP-hard which is certainly not desirable. In this paper we present a new technique that uses IRT to estimate procedural knowledge in a problem-solving environment. To this end, problem solution path is ideally modeled as a directed graph in which nodes are states of the problem and edges are transitions between states.

## 2 Procedural Assessment through Item Response Theory

IRT [4] is one of the best-known strategies for declarative knowledge assessment. According to how the models update the inferred student knowledge in terms of their response IRT-based models could be [5]: dichotomous and polytomous models. Dichotomous models consider only two possible scores, i.e. either correct or incorrect. A characteristic curve, called *Item Characteristic Curve* (ICC), models each item. This curve expresses the probability that student with a certain knowledge level will answer the item correctly; polytomous models have a characteristic curve per choice in an item called *Operating Characteristic Curve* (OCC) [6], which expresses the probability that a student with a certain knowledge level will select this answer [7].

In our approach, each problem is internally modeled as a directed graph the nodes of which are states of the exercises and arcs are the transitions between states. Figure 1 shows an example of a possible graph that represents the addition of  $\frac{9}{14}$ . In this figure, we can see that there is more than one path to reach the correct solution  $\frac{33}{35}$ .

To apply the IRT to estimate the student's knowledge when applying procedural skills, we have made an analogy between problem solving and testing. To do this matching we understand each state of problem solving as an item in testing, and each possible next step from a state to another, as choices of the item. Accordingly, the process of solving a problem by a student could be considered as a branched test where he/she is answering items about the procedures applied to solve this problem. Let us consider the node at the top of the graph presented in Figure 1. This node could be understood as an item such as *How can we go on one step?*, and choices of this item could each be arcs for other nodes of the graph. While solving a problem, there are states where a student could make more than one decision, and therefore, we cannot consider the correctness of a step in isolation. There is usually more than just one way to reach a solution and, generally, they are not all equally good (e.g. adding fractions by multiplying denominators is not as correct as using the least common denominator but it is not incorrect), and even other steps could lead to incorrect solutions. We have chosen an IRT-based polytomous model to assess procedural knowledge since each step could provide relevant information for the assessment. Therefore, each step in problem solving will have its own characteristic curve, like options in polytomous traditional models. The student model is updated during problem solving by the product of the characteristic curves corresponding to the steps he/she is following. In our model, each characteristic curve is called a *Step Characteristic Curve* (SCC). That means that each edge of the problem graph has a characteristic curve associated with it and while a student is "navigating" through the graph, his/her knowledge estimates are updated by means of IRT using the SCC associated to each arc he/she is navigating through.

In this work, for modeling the SCCs we have used the proposal by Thissen and Steinberg for multiple-choice items [8]. The formula of each observable category is shown below, where  $X_i$  represents the item and  $h$  the response selected in this item:

$$P(X_i = h|\theta) = \frac{e^{a_h\theta+c_h}}{\sum_{k=0}^{m_i} e^{a_k\theta+c_k}} + d_h \frac{e^{a_0\theta+c_0}}{\sum_{k=0}^{m_i} e^{a_k\theta+c_k}} \quad (1)$$



Fig. 1. Example of an exercise graph.

### 3 Experimentation

We have conducted an experiment in order to determine whether or not our approach could be useful for procedural knowledge assessment. The experiment was carried out in January 2013 with a sample of 23 undergraduate students from a course of Project Management taught in the ninth semester of the M.Sc. in Computer Science at the University of Málaga (Spain). These students were previously instructed in project investment with a two-hour lecture and a two-hour training problem solving session. Our hypothesis is that the results of measuring the performance of students while solving a problem with our proposal should be similar to those results obtained through a test with items asking about procedural principles.

The experiment was conducted in a two-hour laboratory session and comprised two different phases. Firstly, the students had to solve two project investment problems. Secondly, a test was posed, the goal of which was to assess the same procedural skills evaluated through the test. The two problems posed to students required calculus of some financial indexes in the project investment domain. The test had 15 multiple choice items related to the calculation of these indexes focusing on procedural steps needed to reach the correct result.

We calculated the student's procedural knowledge from the evidence obtained in the two different ways, i.e. in the test and in the problems applying our proposal. Both calibration and assessment were made using the model explained in Section 2 for both the test and our model. Once the calibration and the assessment of test and problem solving were complete, we compared the results of both techniques. We have accomplished correlation tests in order to verify if both estimations were similar. Results show that the correlation between test results and our approach results were different to 0 with 90% confidence obtaining a correlation coefficient of 0.3435.

## 4 Conclusions and Future Work

In this paper, we have presented a new approach for assessing procedural knowledge. This proposal applies the IRT (commonly used to assess declarative knowledge) to assess procedural knowledge in problem solving environments. The main idea behind this proposal is to map a problem with a test. As a result, the process of solving a problem can be seen as a graph the node of which are states and its arc the result of applying procedural knowledge which also leads to new states. In our model each state can be modeled with an item, the choices of which are the set of possible transitions from the source state to another new state, obtaining as a result, what in the testing literature is known as a branched test. We have conducted an experiment in order to explore the performance of our model. To this end, we have constructed a test in which items were focused on procedural skills and we have compared its results with those obtained using our model. The evidence suggests that our model can be used for assessment purposes in a problem solving environment and, consequently, as a tool for updating the student model.

**Acknowledgements:** This work is part of DEDALO project which is financed by the Andalusian Regional Ministry of Science, Innovation and Enterprise (P09-TIC-5105). This work is financed by Universidad de Málaga. Campus de Excelencia Internacional Andalucía Tech

## 5 References

1. Bloom, B.S.: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Leadership* 1 (1984) 4-17.
2. Rittle-Johnson, B., Koedinger, K.R.: Designing Knowledge Scaffolds to Support Mathematical Problem Solving. *Cognition and Instruction* 23, no. 3: 313-349. Routledge (2005).
3. Bisanz, J., LeFevre, J.: Strategic and nonstrategic processing in the development of mathematical cognition. In Bjorklund, D., ed.: *Children's strategies: Contemporary views of cognitive development*, Hillsdale, NJ, Lawrence Erlbaum Associates, Inc. (1990) 213-244.
4. Embretson, S.E., Reise, S.P.: *Item response theory for psychologists*. Lawrence Erlbaum, Mahwah (2000).
5. Guzmán, E., Conejo, R., de-la Cruz, J.L.P.: Adaptive testing for hierarchical student models. *User Model. User-Adapt. Interact.* 17 (2007) 119-157.
6. Dodd, B.G., De Ayala, R.J., Koch, W.R.: Computerized adaptive testing with polytomous items. *Applied Psychological Measurement* 19 (1995) 5-22.
7. Guzmán, E., Conejo, R.: A model for student knowledge diagnosis through adaptive testing. In: *In Proceedings of 7th International Conference Intelligent Tutoring Systems, ITS2004 Brasil*, Springer-Verlag (2004) 12-21.
8. Thissen, D., Steinberg, L.: A response model for multiple choice items. *Psychometrika* 49 (1984) 501-519.