# Measuring Misconceptions Through Item Response Theory

Eduardo Guzmán(✉) and Ricardo Conejo

Dpto. Lenguajes y Ciencias de la Computación, E.T.S.I. Informática,
Universidad de Málaga, Bulevar Louis Pasteur, 25. Campus de Teatinos, 29071 Málaga, Spain
{guzman,conejo}@lcc.uma.es

**Abstract.** In this paper we propose an assessment model to measure both student knowledge and misconceptions through testing. For this purpose we use a well-founded psychometric theory, i.e. the Item Response Theory (IRT). Our proposal is an extension of our previous work in this field and permits, in the same test, the data-driven evaluation of knowledge and several misconceptions, thereby more efficiently using the evidence provided by the students, while solving a test, to enrich student perturbation models.

**Keywords:** Misconceptions · Assessment · Student modeling · Item Response Theory

## 1 Introduction

A tutoring system uses the information stored in a student model to tailor the way it interacts with a student [1]. A precise student model should contain not only information about learning (overlay modeling), but also data about errors made by the learners and the misconceptions they may have (i.e. perturbation models [2]). *Misconceptions refer to ideas that learners have incorporated into their cognitive.*

Testing is perhaps the most extended strategy for assessment. Among the underlying techniques for computing a student's state of learning, the *Item Response Theory* (IRT) is the most widely used when accurate and invariant diagnostic measures are required. In IRT [3] diagnostics are made in terms of the evidence provided by the students through their performance in a set of *items* (e.g. test questions). IRT is based on two principles: a) The student's performance in a test can be explained by means of a single trait (generally, in educational domains, the knowledge level), which can be measured as an unknown numerical value. b) The performance of a student with an estimated trait level answering an item *i* can be probabilistically predicted and modeled by means of a function called *Item Characteristic Curve* (ICC). It expresses the probability that a student with certain trait level $\theta$ has to answer the item correctly. The greater the student's trait level, the higher the probability of them answering the item correctly.

However IRT determines the student's score by identifying his/her location along a single proficiency continuum and therefore does not provide sufficient data to


C. Conati et al. (Eds.): AIED 2015, LNAI 9112, pp. 608–611, 2015.
DOI: 10.1007/978-3-319-19773-9_73

enhance instruction and learning [4]. In this paper we present a model based on the hypothesis of that certain incorrect choices of test questions can be used to infer students' misconceptions. Accordingly, after a testing session, the student model could be updated not only with information about his/her knowledge, but also with data about his/her state of misconception. This proposal is also an extension of our student knowledge diagnosis model [5] currently implemented in, Siette (www.siette.org), our web-based system for automatic assessment [6].

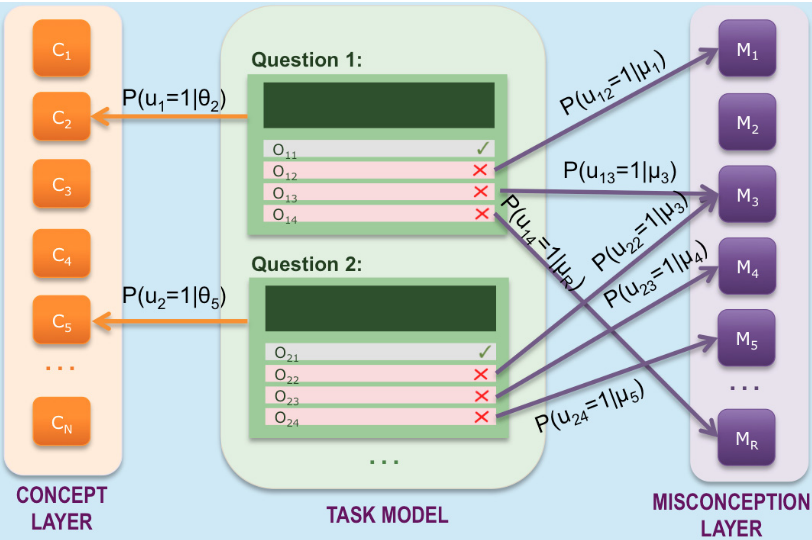


**Fig. 1.** Relationship between tasks (questions), concepts and misconceptions

## 2 A Model for Assessing Misconceptions Through IRT

Our model is framed under the *Evidence-Centered Design* (ECD) proposal, i.e. a guideline for designing, producing and delivering educational assessments [7]. In accordance with that idea, our proposal is composed of the following three submodels: (1) *Student model*, is formed by a *concept layer*, consisting of the set of domain concepts $C_1,\ldots,C_N$, and a *misconception layer* consisting of a set of misconceptions $M_1,\ldots,M_R$. The set of concepts and misconceptions are measured through probability distributions, which relate the student's level in the concept or misconception, and the probability of having it. (2) *Task model*, which consists of the elements through which evidence of knowledge can be captured. It is formed, therefore, by assessment activities (e.g. exercises, problems or test questions). (3) *Evidence model*: It is the connection between the two previous models. It is used thus to transform the raw observations about the student's performance into updates in his/her student model. This process can be carried out thanks to the underlying relationship of tasks with concepts and with misconceptions (see Fig. 1). The transformation is done according to the response model, explained below.

Incorrect responses to items can provide evidence of this misconception. Let us consider, for instance, the algebra domain in which a student is solving questions in a test about fractions. If the student in question does not know how to add fractions correctly, he/she may think, for example, that the fraction resulting from adding two fractions has a numerator equal to the sum of the numerators, and that the denominator is also the addition of the denominators. If, in a test there are several questions involving the adding up of two fractions, and these questions have an option where this addition is calculated wrongly in the way that the student misunderstands, and therefore this will be the response chosen. Fig. 1 summarizes this hypothesis, graphically. In the figure, the task model has been simplified to a set of two questions with four choices. Each question $Q_i$ is linked to one concept $C_j$ modeling the fact that that item can be used to assess $C_j$. In the figure, $Q_1$ assesses $C_2$ and $Q_2$ assesses $C_5$. Regarding the relationship between questions and misconceptions, $Q_i$ could be related to more than one misconception. For instance, $Q_1$ can provide evidence about $M_1$, $M_3$ and $M_R$. In the figure, if a student holds $M_3$, when posed $Q_2$, she will tend to select choice $o_{22}$.

In our proposal, relationships between questions and concepts and between questions and misconceptions is modeled by characteristic curves. The first relationship is the classical one used in IRT and is represented by the ICCs. For example, ICC $P(u_i=1|\theta_2)$ relates the performance of students in $Q_1$ with their knowledge in $C_2$. However, we add a new type of characteristic curve, the *Misconception Characteristic Curve* (MCC), $P(o_{ij}=1|\ \mu_k)$ which models the probability of selecting the *i-th* choice of the *j-th* item, that is, $o_{ij}=1$, given the student level $\mu_k$ in misconception $M_k$. This data-driven curve can be modeled with the same functions as the ICCs, since it is also an increasing monotone function (the greater the misconception level, the higher the probability of selecting that item choice).

Our assessment or diagnostic algorithm is an extension of our previous work summarized in [5]. Let us assume a student is taking a test. The assessment procedure will consist of the following steps:

1. For each concept $C_i$ involved in the test, an equiprobable probability distribution $P(\theta_i)$ will be initialized. Analogously, for each misconception $M_i$ involved in the test another equiprobable probability distribution, $P(\mu_i)$, will be also initialized.
2. Each time the student answers a question $Q_i$ choosing the j-th choice $c_{ij}$:
   2.1. If $Q_i$ assesses $C_k$, the probability distribution $P(\theta_k)$ will be updated with the ICC, $P(u_i=1|\theta_k)$, if the answer is correct. Otherwise, the opposite curve to ICC, i.e. $1-P(u_i=1|\theta_k)$, will be used to update $P(\theta_k)$:

$$\mathrm{P}(\theta_k) = P(\mathrm{u_i}=1|\theta_k)^{u_i}(1-P(\mathrm{u_i}=1|\theta_k))^{(1-u_i)}P(\theta_k) \qquad (1)$$

   2.2. If $Q_i$ is related to one or more misconceptions, for each one of them, its probability distribution will be updated. Let $M_r$ be one of these misconceptions. The probability distribution $P(\mu_r)$ will be updated with the MCC, $P(o_{ij}=1|\ \mu_r)$, if this misconception is linked with choice $o_{ij}$. Otherwise the opposite curve to MCC, $1-P(o_{ij}=1|\ \mu_r)$, will be used:

$$\mathrm{P}(\mu_k) = P\,(\mathrm{o_{ij}} = 1|\mu_k)^{o_{ij}}(1 - P\,(\mathrm{o_{ij}} = 1|\mu_k))^{(1-o_{ij})}P(\mu_k) \tag{2}$$

3. Step 2 will be repeated for each question posed to the student.

The process of assessment will give, as a result, a set of concept or misconception probability distributions. The student's level in that concept or misconception can be computed directly from its probability distribution. The Bayesian MAP or EAP estimators can be used to infer this value.

## 3 Conclusions

This paper has presented a new approach for assessing misconceptions through testing. Historically, tests have been used to measure knowledge. However, with our model, assessment information provided by a test can be optimized by including estimates about certain misconceptions. Estimates about knowledge and misconceptions inferred by our model are invariant and independent of the test thanks to the underlying IRT-based model we use. The main shortcoming of our proposal is related to the test elicitation. The construction of tests with incorrect choices targeting misconceptions is a time consuming task and requires certain experience and an extra effort from tutors. However we think that this model in combination with the inference technique described in [9] could leverage this process. The synergy between both techniques could be useful to produce accurate student models. Moreover, we would like to mention that this proposal will be integrated into Siette in the near future.

**Acknowledgements**: This work is part of DEDALO project which has been financed by the Andalusian Regional Government (P09-TIC-5105).

## References

1. Millán, E., Loboda, T., Pérez-de-la-Cruz, J.L.: Bayesian networks for student model engineering. Computers & Education **55**(4), 1663–1683 (2010)
2. Brown, J.S., Burton, R.R.: Diagnostic models for procedural bugs in basic mathematics skills. Cognitive Science **2**, 155–192 (1978)
3. Hambleton, R.K., Swaminathan, H., Rogers, H.J.: Fundamentals of Item Response Theory. Sage, Newbury (1991)
4. de la Torre, J., Minchen, N.: Cognitively diagnostic assessments and the cognitive diagnosis model framework. Psicología Educativa **20**(2), 89–97 (2014)
5. Guzmán, E., Conejo, R., Pérez-de-la-Cruz, J.L.: Adaptive Testing for Hierarchical Student Models. User Modeling and User-Adapted Interaction **17**, 119–157 (2007)
6. Conejo, R., Guzmán, E., Millán, E., Trella, M., Pérez-De-La-Cruz, J.L., Ríos, A.: SIETTE: A web-based tool for adaptive testing. IJAIED **14**(1), 29–61 (2004)
7. Mislevy, R.J., Almond, R.G., Lukas, J.F.: A Brief Introduction to Evidence-Centered Design, CSE Report 632, National Center for Research on Evaluation, Standards and Student Testing (CRESST), May 2004
8. Guzmán, E., Conejo, R., Gálvez, J.: A data-driven technique for misconception elicitation. In: De Bra, P., Kobsa, A., Chin, D. (eds.) UMAP 2010. LNCS, vol. 6075, pp. 243–254. Springer, Heidelberg (2010)