# Towards a computational theory of learning in an adaptive testing environment

Ricardo Conejo Eduardo Guzman
Jose-Luis Perez-de-la-Cruz
*Departamento de Lenguajes y Ciencias de la Computación. Universidad de Málaga, Spain.*
{conejo,guzman,perez}@lcc.uma.es

**Abstract.** This paper presents a theoretical model of the learning effects of hints and feedback provided to the student while taking a test. We analyze the properties of feedback and state some formal axioms that every model of feedback must satisfy.

## 1 Tutoring while testing: hints and feedback

Based on Item Response Theory, our group has developed and implemented the system SIETTE [1]. SIETTE can be used not just for assessment, but also for instruction. This is usually achieved by means of *hints* and *feedback*. In the following we will assume that the student is challenged to solve a *test* and that the system can provide adaptively *hints* and/or *feedback*. First of all lets define which is the intended meaning of these terms.

Item. We use this term denote a question or exercise posed to a student. An item consists in a multiple choice question that is the conjunction of a *stem* and a set of possible *answers* where only one is correct. A *test* is a sequence of items.

Hint. A hint is an additional piece of information that is presented to the student after posing a question and before he answers it. Hints may provide an explanation of the stem, or some indication to reject one or more of the possible answers. Given an item, none, one, or more than one hint elements may be available.

Feedback. In this paper, feedback is defined as an additional piece of information that is presented to the student after he tries to answer an item. Feedback is usually given to correct a wrong answer (negative feedback), but it can be also given to reinforce a correct answer (positive feedback). Given an item and a possible answer, none, one, or more than one feedback elements may be available (normally just one for each pair item-answer).

The use of hints as they have been defined above do not modify the nature of the testing process. In fact, if hints are just explanations of the stem of the item, or warn the examinee that some options of the tests are not right, then they do not make any significant difference in the essence of the test. In this case it can be supposed that the examinee's knowledge is not changing, but the question has became easier, and so she might be able to solve it and the conjunction of the current item plus the hint could be considered as a new item. This new (virtual) item can be treated and measured in the same way that all other items in the test. A new $ICC$ can be assigned to the new item and its parameters can be estimated in the same way that for all other items in the test. However, both $ICC$'s are not independent. First at all, as we have just mentioned, the use of a hint makes the question *easier*. This condition can be stated in mathematical terms by the following

**Axiom 1.** *Given a question $q$ and a hint $h$, for all knowledge levels $k$, $ICC_q(k) \leq ICC_{q+\{h\}}(k)$, where $ICC_q$ represent the original item characteristic curve and $ICC_{q+\{h\}}$ represent the characteristic curve of the item with the hint.*

Standard procedures can be used to estimate the values of $ICC_{q+h}(k)$ according to student responses. If the examinee uses a combination of hints, the question should become even easier. Thus we arrive to the following

**Axiom 2.** *Given a question $q$, a set of hints $H$ and a hint $h \notin H$, for all knowledge levels $k$, $ICC_{q+H}(k) \leq ICC_{q+H+\{h\}}(k)$.*

If the empirical estimation of the parameters does not satisfy the above axioms, the hint should be rejected (because it behaves as a misleading element). From an ITS point of view, the interesting point is the following: giving two possible hint elements, which one should be applied first?. This question is easy to answer in an adaptive environment, by the application of the classical adaptive mechanism. Given a knowledge estimation $\theta(k)$ for a student, and given two hints $h_1$, $h_2$, with $ICC_{q+\{h_1\}}(k)$ and $ICC_{q+\{h_2\}}(k)$, the best hint to use is the one that makes lower the expected variance of the posterior probability distribution. This mechanism is simple to implement and do not make a substantial modification of the adaptive testing procedure.

## 2 A model of knowledge change

Adaptive testing must solve the question of measuring the examinee's knowledge in a given instant of the exam. When we assume that testing and learning are interleaved, the problem is even more complex. In this section we will define a general model of learning while receiving feedback.

Let $\varphi$ be a feedback element. Let us assume that the level of knowledge of a student can change when $\varphi$ is given. Moreover, let us assume that the change depends on the prior level of knowledge. Let us consider two stochastic variables, $\theta_1$ for the prior level of knowledge and $\theta_2$ for the posterior level of knowledge. Let us denote with $f(k_1, k_2)$ the corresponding density function for the values $k_2$ of $\theta_2$ given that $\theta_1 = k_1$. It is obvious that for each $k_1$

$$\int_0^\infty f(k_1, k_2) dk_2 = 1.$$

Moreover, in our domain, it makes sense to impose $f(k_1, k_2)$ the following condition:

**Axiom 3.** *(Monotonicity). It is impossible that the level of knowledge decreases after the feedback element is given, i. e., if $k_1 > k_2$, then $f(k_1, k_2) = 0$.*

For the values $k_2 \geq k_1$, the probability must reach a maximum (usually, near to $k_1$) and then decrease. That is the meaning of the following

**Axiom 4.** *(Gradual learning). For each fixed prior level of knowledge $k_1$, the function $f(k_1, k_2)$ is unimodal in $k_2$.*

Lets consider the expected value $E(\theta_2)$ of $\theta_2$ for each prior value $k_1$ given by

$$E(\theta_2) = g(k_1) = \int_{k_1}^{\infty} f(k_1, k_2) k_2 dk_2$$

It makes also sense to assume that the same instructional action applied to different levels of knowledge can at most equal them, but can never shuffle their ordering:

**Axiom 5.** *(Stability). The expected value of $\theta_2$ is non-decreasing with the prior level of knowledge $k_1$, i. e., for all $k_1, k_1'$, if $k_1 \leq k_1'$, then $g(k_1) \leq g(k_1')$.*

Finally, let us consider the expected gain of knowledge, given by

$$E(\Delta\theta) = \int_{k_1}^{\infty} f(k_1, k_2)(k_2 - k_1) dk_2 = g(k_1) - k_1.$$

The expected effectiveness of a feedback element is higher in the environment of the student current knowledge. That is, the feedback is more probable to make a significant change in the student knowledge if the student's knowledge is closer to the item's difficulty. Feedback for very easy questions do not make a significant difference in the student knowledge because the concepts involved are already known and feedback to questions that are too difficult is not properly understandable by the student.

**Axiom 6.** *(Vygotskii's law [3]). The expected increment of knowledge is unimodal on the prior level of knowledge $k_1$.*

We will find the maximum of this function for a certain value $k^*$. We hypothesize that this $k^*$ will be in the environment of the item difficulty parameter. Assuming that $f(k_1, k_2)$ is known for every feedback element, the application of an adaptive test is easy. First a question is posed; if the students answers, the posterior probability $p(k)$ of the student knowledge is calculated for every value $k$. Then, if a feedback element $\varphi$ is given to the student, the posterior probabilities just obtained are used as the prior ones in the function $f$, and a new distribution $p'(k')$ of the estimated knowledge level is obtained by

$$p'(k') = \int_{0}^{\infty} p(k) f(k, k') dk'.$$

The adaptive mechanism of the test is not changed and the test can continue. If there are more than one feedback element for a given answer, the one that makes a higher expected improvement in the student's knowledge level should be used. Murray and Arroyo [2] have proposed another operational model that tries to guarantee that scaffolding remains in the Vygotskii's Zone of Proximal Development (ZPD). The main difference with our approach is that we do not consider all feedback elements as having equal effectiveness. In our case the IRT mechanism can keep by itself the difficulty of the next item in the ZPD, and what we measure is the real effect of different scaffoldings.

## References


[1] R. Conejo, E. Millán, J. L. Perez-de-la-Cruz and M. Trella, SIETTE: A Web-Based Tool for Adaptive Teaching. International Journal of Artificial Intelligence in Education (2003) (to appear).

[2] Murray, T. and Arroyo, I., Toward Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems. In: Proceedings of ITS'2002, Biarritz (2002).

[3] Vygotskii, L. S., The Vygotskii Reader, Blackwell, Cambridge, Mass, (1994).