Using the Item Response Theory (IRT) for Educational Evaluation Through Games

Marcelo Henrique Euzébio Batista, FACENSA College, Gravatai, Brazil
Jorge Luis Victória Barbosa, University of Vale do Rio dos Sinos (UNISINOS), Sao Leopoldo, Brazil
João Elison da Rosa Tavares, FACENSA College, Brazil
Jonathan Luis Hackenhaar, University of Vale do Rio dos Sinos (UNISINOS), Sao Leopoldo, Brazil

ABSTRACT

This article shows the application of Item Response Theory (IRT) for educational evaluation using games. The article proposes a computational model to create user profiles, called Psychometric Profile Generator (PPG). PPG uses the IRT mathematical model for exploring the levels of skills and behaviors in the form of items and/or stimuli. The model was integrated with an educational game. The game was created considering a test applied by the Brazilian Government to evaluate students in national level, called Brazil Exam. The integration was used to evaluate the model in an educational scenario involving one hundred and thirteen students with an average age of 11 years old of a school in the south of Brazil. The results show PPG conducts an accurate evaluation because it considers not only the number of questions correctly answered but the proportional difficulty of each question in the evaluated group.

Keywords: Educational Game, Item Response Theory, Psychometric Profile Generator (PPG), Psychometrics, User Profile

INTRODUCTION

Rossous (2007) affirms the information and communication technologies have assumed a critically important role in almost every facet of our society. He also highlights the competence in using computers is not only an advantage, but also often a necessity.

When the student is evaluated without a methodology, ends up getting a concept note or other classification which had no sense to improve the knowledge or to enhance the review of what was not understood to be rethought and improved. An evaluation methodology should provide subsidies, favoring the development and evolution, in order to make the most of the student.

DOI: 10.4018/jicte.2013070103
In last years, the Brazilian government has implemented a form of evaluation at national level. Once a year, a test is applied to students in public education. This test is called Brazil Exam (in Portuguese “Prova Brasil”) and its main goal is to get data on the ability and aptitude of the students. Currently, the Brazil Exam is implementing the Item Response Theory (IRT) (Hambleton & Swaminathan, 1985; Rainer & Miller, 1996), that consists of a mathematical model which exploits the level of skill or learner behavior. According to Zlomke (2009), the IRT may be a useful framework for examining the reliability and latent traits of the measures assessed.

Based on the IRT, this article proposes the PPG (Psychometric Profile Generator) model. The model is composed by different evaluation environments and prospect profiles. In order to assess the PPG model, we constructed an educational game based on it. The game was used in a case study involving one hundred and thirteen students in the 5th grade of basic education with average age of 11 years old.

The article is divided into seven sections. The next section presents the psychometrics and mathematical model used. The following section describes the computational model proposed and its operation. Afterwards we present the case study based on the educational game. Next we discuss the approach to the results of the case study. Followed by a section that compares the PPG with three related works. Finally, the last section approaches the final considerations.

The Psychometrics

Psychometrics (Arigbabu, 2009; Peterson & Johannson, 2006; Raykov & Marcoulides, 2011) is a sub area of psychology that uses Statistics to study psychological phenomena. The cycle starts with the definition by the specialist (Psychologist or Educator) of which behavior or skill will be measured. The next step is to construct items and/or stimuli that actually represent the specific behavior or skill defined in the previous step (Santos & Guedes, 2005).

In this sense, it is necessary to establish the dimensionality of the behavior or skill, and define their constitutively and operationally, that will be detect through questions, operations it in tasks, items of skills and/or behavioral stimuli (Anastasi & Urbina, 2000).

However, research on this topic is scarce and separate norms might be needed. (Hedemann et al., 2010). Through mathematical model known as IRT (Hambleton & Smawinathan, 1985) and represented in Equation 1, it is possible to estimate a profile of each subject and collect information about each item of a test by checking the latent trait $P_i(\theta)$ and matching items for each subject according to the objectives proposed.

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \quad (i = 1, 2, \ldots, n)$$

In Equation 1, $P_i(\theta)$ is the probability of a subject correctly answers an item or stimulus. The constants $a_i$, $b_i$ and $c_i$ are already prospected values by IRT. In turn, $n$ is the number of evaluated items applied and the constant $e$ is the number of Euler (2.7182818). Finally, the constant $D$ has a value of 1.7. The model for IRT is represented graphically by the function in the form of a characteristic curve related to the item, as shown in Figure 1.

Estimation of the parameters is done using procedures based on methods of maximum likelihood. The equations estimated by these methods cannot be solved analytically, so they must be solved by interactive numerical procedures. The interactive methods commonly used in IRT are the algorithm “Newton-Raphson” and the method “Scoring” Fisher, among others like the methods of squaring algorithm and “Hope and Maximization”.

The function of Characteristic Curve of Item is shown in the Figure 1, with three parameters measured. The parameter “a” is represented by the curve at the inflection point where the curve crosses the line representing the probability of 50%. The “b” parameter measures the difficulty of the stimulus being represented by the distance along the lines of X (abscissa), which corresponds to the determined point by the perpendicular that comes from
Related Content

A Model for an Adaptive e-Learning Hypermedia System
www.irma-international.org/article/a-model-for-an-adaptive-e-learning-hypermedia-system/99627/

Validation of Learning Effort Algorithm for Real-Time Non-Interfering Based Diagnostic Technique
Pi-Shan Hsu and Te-Jeng Chang (2011). International Journal of Distance Education Technologies (pp. 31-44).
www.irma-international.org/article/validation-learning-effort-algorithm-real/55797/

Fostering Successful Learning Communities to Meet the Diverse Needs of University Students by Creating Brain Based Online Learning Environments
www.irma-international.org/article/fostering-successful-learning-communities-meet/37516/

Data Flow Diagrams vs. Use Cases – Student Perceptions
www.irma-international.org/article/data-flow-diagrams-use-cases/2310/

Delphi and NGT for Consensus Building E-Research
www.irma-international.org/chapter/delphi-ngt-consensus-building-research/11806/