Estimating Programming Skills with Combined M-ERS and ELO Multidimensional Models

1st Fabiana Zaffalon
PPG Sciences Education
Federal University of Rio Grande
Rio Grande, Brazil
fabinhazaffalon@gmail.com

2nd André Prisco
PPG Sciences Education
Federal University of Rio Grande
Rio Grande, Brazil

3rd Ricardo de Souza
PPG Computational Modeling
Federal University of Rio Grande
Rio Grande, Brazil

4th Jean Luca Bez
Integrated Regional University
Erechim, Brazil

5th Neilor Tonin
Integrated Regional University
Erechim, Brazil

6th Rafael Penna
Center of Computational Sciences
Federal University of Rio Grande
Rio Grande, Brazil

7st Silvia Botelho
PPG Sciences Education
Federal University of Rio Grande
Rio Grande, Brazil

***Abstract*—This complete article, from the research to practice category, presents an experiment carried out combining two models used to evaluate student skills, ELO Multidimensional and M-ERS. The objective of this experiment is to estimate and map the history of their multiple skills, in that way it was carried out incorporating the characteristic of the Multidimensional ELO - to track the history of multiple skills, and M-ERS - to estimating multiple skills that can be compensatory. To validate the experiment, we used a database composed of user submissions from an Online Judge platform from Brazil. Through the experiment results obtained, we concluded that for online programming problems platforms, the combination of both models proved to be satisfactory, through it was possible to map and observe the evolution of student's multiple skills.**

***Index Terms*—Skill, sub-skill, difficulty, IRT, MIRT, ELO, M-ERS.**

I. Introduction

In recent years, the most prominent educational assessment methods are those that aim to present accurate data on the construction of competences or skills acquired by students, not by a punctual evaluation, but by an assessment that tracks the student's progress over time and that takes into account the various skills involved in solving problems, whether from presential or distance students.

Recommendation systems incorporated into education, presential or at a distance, are important resources because they offer easy access to study materials and assist students in the search and selection of content focused on their profile [2]. Such systems filter the contents to be recommended through the collection and analysis of the student's usage profile, thus indicating more relevant materials [2].

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) – Finance Code 001.

Also, for *educational recommendation systems* that propose exercise resolutions, a model for estimating potential strengths and weaknesses in student's skill sets can provide a more suitable indication, avoiding the frustration of problems that are too difficult, or the comfort of the too easy ones.

There are platforms for students in the field of computing that offer materials and exercises for computer programming practice, which encompasses a set of skills such as logical reasoning and mathematics, and also others not directly related to computer science curricula, text interpretation as an example [7], [17]. Therefore, it is important to evaluate this type of problem, as it is possible to more accurately identify which skill needs to be developed on a student.

There are methods that estimate student's skills, and two prominent models are the Item Response Theory (IRT) [1] and the ELO Rating System (ERS) [5]. The IRT is a set of mathematical models that seeks to estimate the probability of an individual answering a question about an item correctly, establishing parameters that represent the relationship between the difficulty of an item and the ability of an student to answer it properly [1], [3].

IRT presents compensatory multidimensional models that contemplate cases in which more than one skill is required for the student to correctly answers an item, and that the low skills can be compensated for by higher skills [1], [3].

On the other hand, ELO, widely used for evaluation in international chess rankings, aims to classify players through their game histories, using a statistical classification that calculates skill for competitors. The model deals with a player's relationship with another player. However, in education, this combination happens between the student and the problem with which he interacts [14].

There is also the Multidimensional ELO model where each dimension is a skill developed to solve programming problems [13], however, this model assumes that the relationship between the student and the learning objects comprises a set of independent, orthogonal skills.

According to [12] what differentiates the two approaches is the procedure of estimating the parameters and in their basic assumptions: IRT assumes that the student's ability is constant, while ELO was proposed to model changes in skill level.

The E-MERS approach proposed by [10] uses a compensatory MIRT model and the ELO model to estimate student's skills in adaptive learning systems, assuming that an item may involve more than one skill and that a low skill can be compensated for a higher one.

Considering that the programming exercises involve skills that are interdependent, in this work we performed an experiment combining characteristics of the ELO Multidimensional model and the M-ERS model. The objective of this work is to find a model that simultaneously estimates and updates the skills of students in order to allow the history mapping of the multiple skills that involve programming problems.

To validate the model, we used a database made available by an Online Judge do Brazil platform, composed of a sample of 10.000 submissions, organized in chronological order, of 99 users and 99 problems.

In the following sections, we present the theoretical background that supported the research, the methodology used for the execution of the experiment, analysis of the results obtained and, finally, the conclusions.

II. Item Response Theory

Item Response Theory (IRT) is an approach that has been gradually inserted in the education space for being considered an important instrument in the quantitative process of educational evaluation, as it allows the measurement of characteristics that otherwise would be difficult to be evaluated directly (skills) [1], [3]. To calculate the estimated skill, also called *Theta* (θ), the IRT is based on statistical methods and mathematical models that consider not only the responses of individuals but also the properties of the items [6].

Considering one or more item parameters, that will be presented in the following paragraphs, the greater the individual's skill, the greater his probability to give a correct answer. There are several proposed IRT models, which depend on three factors: a) nature of the item - dichotomous or non-dichotomous; b) number of populations involved; c) amount of skills to be measured - one or more [1], [3].

For dichotomous items, there are 3 models that differ from each other by the number of parameters used to describe item [1], [3]:

- Rasch model (1 parameter): considers only the difficulty of the item.
- 2PL model (2 parameters): considers the difficulty of the item and its discrimination.
- 3PL model (3 parameters): considers difficulty, discrimination and probability a randomly given right answer (chance of individuals with low skill to rightly answer difficult items).

The 3PL Model, whose equation is presented in (1), describes the probability of the individual with the ability θ to correctly answering the item j, taking into account the discrimination of item a, difficulty of item b and the chance of a randomly right answer by chance c [1].

$$P(\theta) = c_j + (1 - c_j)\frac{1}{1 + e^{-a_j(\theta - b_j)}} \tag{1}$$

The item discrimination parameter a_j indicates how much an item distinguishes individuals with different skill levels, when we understand that the same test will be applied to different individuals with different skills [18].

The difficulty of item b_j occurs on the same scale as the skill, that is, the probability of an individual to correctly answer an item is given by the skill necessary for it [9], [15].

The random parameter c_j is the probability that an individual with low skill to rightly answer an item by chance. In cases where random right answers is not possible, like essay tests, parameter c assumes zero value [3], [9].

These models cited consider that the test is a one-dimensional instrument that implies the existence or predominance of only one skill influencing the answers given. However, such an argument does not apply in many practical situations, as an example we can cite a math test that requires text interpretation even before requiring mathematical development, so in this case we would be dealing with a two-dimensional test, as it requires two skills [8].

Researches have showed that the Multidimensional IRT model (MIRT) is better suited to real data than one-dimensional models, as the situation exemplified above is commonplace in Education, in which subject's responses are determined by more than one skill at the same time [11].

MIRT models can be separated into two classes: compensatory and non-compensatory. A model is said to be compensatory when the probability of an item to be successfully answered is maintained or increased even if the value of one of the skills is low, which is compensated by a high value in another skill [8]. We can also use the example of proficiency in mathematics where an individual who has low proficiency in trigonometry, but high proficiency in algebra and text interpretation, may have an increased probability of getting the question right, that is, one skill pays off the other [4].

The compensatory model of MIRT is given by (2) [16]:

$$P_{ij} = P(Y_i = 1) = \frac{1}{1 + exp\left[-\sum_{k=1}^{p} a_{ki}\theta_j + b_i\right]} \tag{2}$$

Where a_i is a vector that contains the discrimination parameter and b is a scalar indicating the item's difficulty.

III. ELO classification system

Initially, the ELO classification system (ERS) was proposed to analyze and classify the performance of chess players [5].

Through statistical methods, each player receives an initial rating (θ_i) and, as they participate in the games, this rating is updated according to the results. The model works as a function of the expectation and the result, the expected probability that the player wins the game is given by the logistic function in relation to the difference in the estimated ratings, as (3) [5]:

$$P(R_{ij} = 1) = \frac{1}{1 + 10^{\frac{\theta_j - \theta_i}{400}}} \tag{3}$$

Where R = {0,1} is the result of a game: 1 (victory) and 0 (defeat). Given a game between player i and player j, with an ELO (θ_i) and (θ_j), respectively, at the end of the match new ELOs are calculated according to the expectations of the results, the previous ELOs and a constant k. The higher the k, the greater the change in ELO, according to (4) [5]:

$$\theta_i = \theta_i + k(R_{ij} - P(R_{ij} = 1)) \tag{4}$$

When applied in education, ELO establishes that a student is considered a player and the problem is considered his opponent [12]. Thus, assuming that an answer to a problem has a relationship between the student and the problem, ELO will be estimating the student's ability and the difficulty of the items. The estimation takes place continuously, as the classification is updated at the end of each resolution [12].

The use of the ELO classification system offers many advantages: it is a simple system, easy to implement in educational applications; it has only a small number of parameters that need to be defined; can be easily used in an online environment; and it also provides good performance for more complex models [12]. However, it is intended to track only a single skill [10].

IV. Related Works

In this section we briefly describe two models that have the purpose of estimating the student's multiple skills, and that are used in the experiments carried out in this work.

A. Multidimensional ELO

The objective is to provide means to combine learning objects and students in a recommendation environment with programming problems, considering that each student has some more developed skills and others that need to be improved.

ELO's original approach to education considers students and learning objects to be a one-dimensional model, assuming that the theme of the proposed challenges can be reduced to a single skill level. However, in the case of programming exercises, it is assumed that they are formed by different levels of maturity and skills.

Therefore, [13] assumes that the relationship between the student and the learning objects comprises a set of skills that can be relatively independent or orthogonal. In this context, a learning object may require more advanced levels in one skill and more basic levels in others, being possible for a problem to completely ignore one or more skills int the set. Likewise, a student can have different levels for each skill.

In the classic model, ELO is a scalar value for each student and for each learning object. The extended model of the ELO, in order to enable it to be multidimensional, considers that each dimension is a skill or concept that the student must have at some level. Each student has their skills represented by (5):

$$\overrightarrow{\theta_s} = (S_1, S_2, ..., S_N) \tag{5}$$

Where S_i is the ELO in the i skill. Each learning object l has its requirement level represented by $\overrightarrow{\sigma}$ and by relevance $\overrightarrow{m_l}$, where n is the number of learning objects. The set of learning objects can be represented by (6):

$$L = \{(\overrightarrow{\sigma_1}, \overrightarrow{m_1}), (\overrightarrow{\sigma_2}, \overrightarrow{m_2}), ..., (\overrightarrow{\sigma_n}, \overrightarrow{m_n})\} \tag{6}$$

Relevance m is a real number, between 0 and 1, which indicates how important a skill is for interaction with a learning object.

Considering the interaction of the student i with the learning object j, the adaptation to the model can be represented by (7). For each s skill:

$$\begin{aligned} \theta_{i_s} &= \theta_{i_s} + m_{j_s} k(R_{ij} - P(R_{ij} = 1)) \\ \sigma_{i_s} &= \sigma_{i_s} + m_{j_s} k(R_{ji} - P(R_{ji} = 1)) \end{aligned} \tag{7}$$

Each interaction is a tuple I = (S, L, R), where R = {0.1}, with 1 indicating that answer was right or 0 indicating that it was wrong.

B. M-ERS Model

The model exposed in [10] presents an approach that incorporates a MIRT model to the ELO model, to track the estimate of the (continuous) skill parameters in adaptive learning systems.

The idea is, instead of assuming a one-dimensional trace of item responses, the approach assumes that a single item can involve more than one skill parameter. In this way, the authors extended the standard ELO, which updates a single parameter, in order to allow a simultaneous update of several skill parameters based on a compensatory MIRT model, as (8) [10].

$$P_{ij} = P(Y_{ij} = 1) = \frac{exp(\sum_{m=1}^{M} \alpha_{jm}\theta_{im} - \beta_j)}{1 + exp(\sum_{m=1}^{M} \alpha_{jm}\theta_{im} - \beta_j)'} \tag{8}$$

The difference between the observed performance and the expected P_{ij} based on the MIRT models is used to update the skill parameters after each item response. P_{ij} within the ELO for the m -th capacity of the individual i is updated according to (9) [10]:

$$\begin{aligned} \hat{\theta}_{im(t)} &= \hat{\theta}_{im(t-1)} + D_{m(t)} K\{Y_{ij(t)} - P_{ij(t)}\} \\ \hat{\beta}_{j(t)} &= \hat{\beta}_{j(t-1)} - D_{m(t)} K\{Y_{ij(t)} - P_{ij(t)}\} \end{aligned} \tag{9}$$

Where $D_{m(t)}$ is a weight to specify whether the m capacity is indicated by the item given in the t -th step. For the skill that is indicated by the item, $D_{m(t)}$ is equal to 1. For the skill that is not indicated by the item, the weight takes values between 0 and 1 [10].

V. Model applied in the experiment

As our focus is on the traceability of student's multiple skills, henceforth referred to as sub-skills, aimed at programming exercises, we conducted an experiment using the two models presented in the previous section.

The equation (3), used in [13] to estimate the student's expected probability of correct response to i to the item j has been replaced by the equation (8) from model M-ERS. We adjusted some parameters of the equation to better fit our data, as (10):

$$P_{ij} = P(Y_{ij} = 1) = \frac{exp(\sum_{m=1}^{M} \alpha_j(\theta_{im} - \beta_j))}{1 + exp(\sum_{m=1}^{M} \alpha_j(\theta_{im} - \beta_j))'} \quad (10)$$

We consider α_j to be the item discrimination j calculated by the IRT 2PL model. θ_{im} is the student's skill i in the m sub-skill. β_j is the difficulty of the item j.

The parameters of student skills and item difficulty are updated with each submission based on (7) of the ELO Multidimensional model with some adjustments, as (11):

$$\begin{aligned} \hat{\theta}_{im(t)} &= \hat{\theta}_{im(t-1)} + \sigma_{jm} K\{Y_{ij} - P_{ij}\} \\ \hat{\beta}_{j(t)} &= \hat{\beta}_{j(t-1)} + \sigma_{jm} K\{Y_{ji} - P_{ji}\} \end{aligned} \quad (11)$$

Where σ_{jm} is the relevance of the m sub-skill of the item j. K is a constant equal to 0.4, according to [19]. In the next section we present the methodology adopted for the application of the experiment.

VI. Methodology

To carry out the research, we used the data provided by an Online Judges platform from Brazil that has a repository of programming problems in which students can solve problems by submitting their solutions in one of the languages accepted by the platform. On this platform, submitted programs are automatically evaluated and students receive feedback. The types of feedback are: accepted (problem accepted without error) or error (compilation, execution, presentation, execution time, time exceeded, communication failure with the server or wrong answer). Teachers and students have incorporated the use of this platform as a educational tool, in addition, the platform is also used for programming competitions.

We used submissions, organized in chronological order, from 99 users of the platform, which totaled a sample of 10.000 submissions in 99 problems. Users chose the problems to be solved and were able to submit several solutions for the same exercise, that way they did not necessarily have solved the same problems.

The sub-skills were selected from the work of [13]:

- Mathematics (sub-skill 1): mastery of algebra and geometry, and knowledge of how to apply such concepts in problem solving.
- Basic (sub-skill 2): involves simpler, sequential problems with operators, conditionals and *loops*.
- Modularization (sub-skill 3): ability to divide a problem into subproblems to enable or facilitate its solution.
- *String* (sub-skill 4): processing of textual data.
- Linear Structures (sub-skill 5): stacks, rows, lists. In some problems, the relevance is high, since the main objective of the problem is to test the student's mastery in some structure.
- Non-linear structures (sub-skill 6): domain and application of graphs and trees.
- Advanced Algorithm (sub-skill 7): covers concepts that go beyond basic programming logic: appropriation of the computational resource concept and the complexity analysis of the algorithm.

The relevance of each sub-skill received real values, between 0 and 1, and the same relationship made by specialists in [13] was used, where each problem received a relevance rate for each of the sub-skills.

To estimate and calibrate the parameters of the problems, we used the IRT model 2PL, which estimates the difficulty and discrimination of the items. We tabulated the data in order to list users, problems and the respective answers (1-correct or 0-incorrect). As users were able to submit several solutions to the same problem, we observed the set of responses given to the same exercise and considered it correct in the cases which any of the submissions were correct.

The difficulty and discrimination estimated by the IRT were attributed to each problem, in addition to the relevance of the sub-skills involved in the problem. Initially, all users were assigned skill and sub-skill values equal to 1, which were subsequently updated for each submission, as well as the difficulty of the problems according to (10) and (11). In the next section, we present and analyze the results obtained.

VII. Results Analysis

After the application of the experiment, we preliminarily analyzed the results obtained. We observe the skill and sub-skill evolution graphs of 3 users chosen at random: 1 student, Fig.1 and Fig.2, that had an increase in the skill value, the second that decreased the value, Fig. 3 and Fig.4, and the third, Fig. 5 and Fig. 6, which presented the final value of the skill close to the initial value.

The original data for the experiment was processes and the models have been applied as described on the previous sections. We have elaborated a series of graphs for a preliminary analysis on the behavior of the generated data. We have observed the skills and sub-skills evolution of 2 users chosen at random: 1 student, Fig. 1 and Fig.2, that had an increase in the skill value and another that decreased the value, Fig. 3 and Fig. 4.

User 1 has 156 submissions of 98 problems with 88 right answers and 67 wrong ones. Through the graph in Fig. 1 we

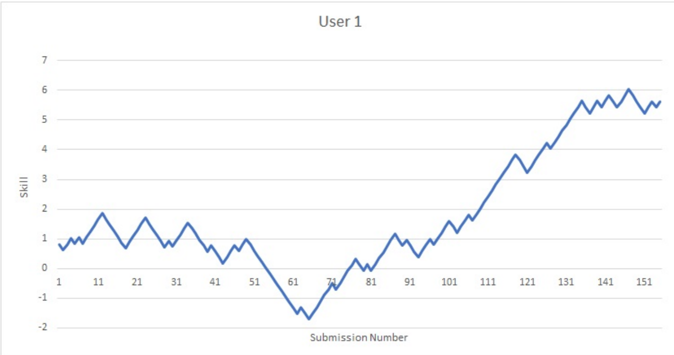

Fig. 1. User 1 skill evolution

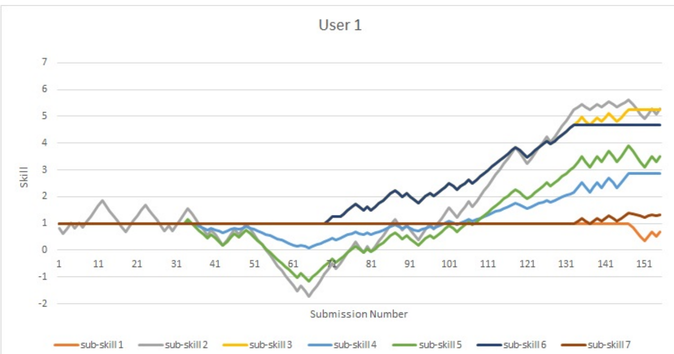

Fig. 2. User 1 sub-skill evolution

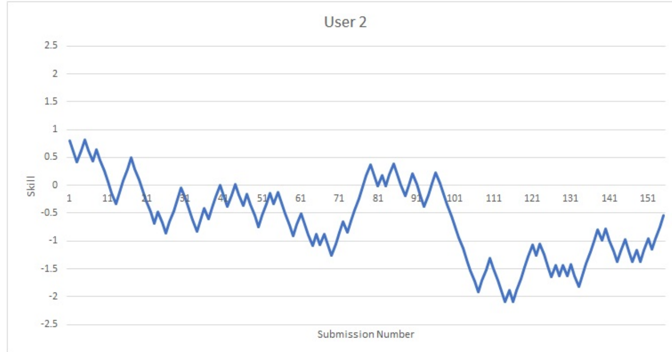

Fig. 3. User 2 skill evolution

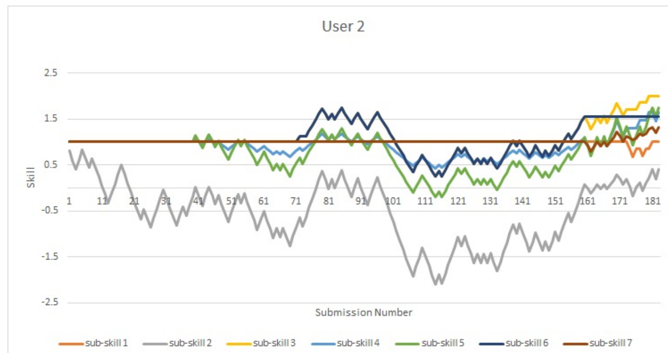

Fig. 4. User 2 sub-skill evolution

observe the positive evolution of the student's general ability. The graph in Fig. 2 illustrates the history of the sub-skills, which vary according to the wrong and correct responses to each submission.

According to the graph in Fig.2, sub-skill 2 (Basic) has relevance in all problems that the student has submitted, these are problems that require the use of programming structures like selection, repetition, etc. The exercises that involve mastery in mathematics (sub-skill 1), for example, started to be solved after submission 148. We noticed that when submitting these problems, the value of this sub-skill decreased, if recovered soon after submitting correct solutions. It is also observed that the exercises that require sub-skill 1, do not require sub-skills 3, 4 and 6.

This user, during his submissions, had an increase in his general skill value (5.62). Most sub-skills oscillated with decreases and increases over time, except for sub-skill 1, which only decreased. So, even with a general skill considered good, that sub-skill needs to be developed. Even though there was a drop in the values of some sub-skills, these did not significantly affect the student's general ability.

User 2 has his skills represented in the graphs of Fig. 3 and Fig. 4. With a drop in the skill value (0.72), he submitted 183 resolutions with 89 right answers and 94 errors. He started to solve the most complex problems, from the 48ª submission. The decrease in the estimated general skill is justified by the drop in highly relevant sub-skills, i.e. the items had high discrimination and difficulty parameters.

The Fig.5 and Fig. 6 illustrate the history of the skill and sub-skills of user 3 who submitted 159 solutions to 96 problems, 88 of which were correct and 91 were wrong.At the end of the submissions, he presented the skill value (0.89) very close to the initial value (1.0). In the first submissions he opted for the easiest problems, gradually increasing the level of difficulty. More complex problems started to be solved from the 71 submission with sub-skill 6 showing high relevance and, from 94 submission, the problems that require sub-skill 7 started to be solved. We can see that this user, despite having decreased his general ability, had an improvement in other skills.

It is possible to notice that the values of the sub-skills are being updated on a greater or lesser scale, according to the stipulated relevance and according to the errors and successes in the solutions submitted.

According to the graphs, it is possible to map the history of all sub-skills, identifying which one the user needs to improve. We were also able to see the type of problems the user solved. As illustrated in the graphs, it is possible to map the history of all sub-skills, identifying which one the user needs to improve.

VIII. Conclusion

Educational online environments offer an opportunity to collect raw data that otherwise would be difficult to obtain,

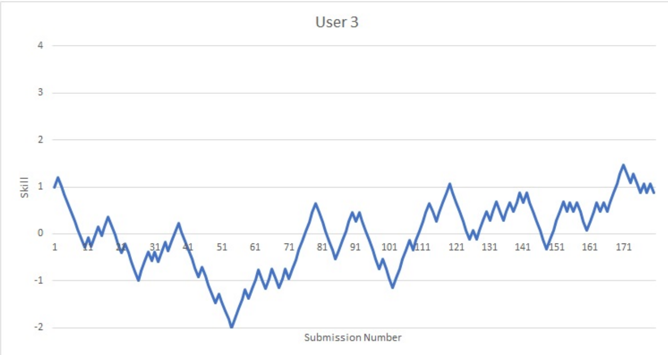

Fig. 5. User 3 skill evolution

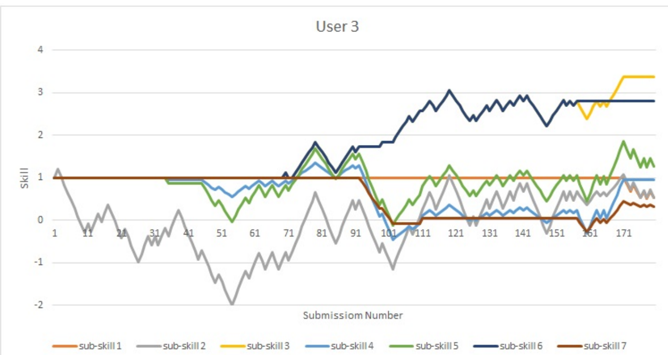

Fig. 6. User 3 sub-skill evolution

computer programming courses also benefit from that characteristic. The generated data can be used in many ways, to estimate skills and sub-skills of students is only one of the possible applications.

In this work, we present an experiment combining two models that estimate multiple skills of students: ELO Multidimensional and M-ERS. Through the experiment carried out, we obtained satisfactory results for our purpose, being able to successfully map the evolution history of student's sub-skills individually.

The resulted data obtained has showed the importance of evaluation of sub-kills to identify potential weakness on the overall skill set of a student, allowing for a punctual educational intervention. Nevertheless, by mapping the data over skills and sub-skills, it was not possible to determine what skill was responsible for every solution failure, for a wrong answer, all skills would be decreased according with its discrimination. In that way, in a submission that involves mathematics and basic skills, even that a student correctly performs all mathematical routines required for the solution, if he misses a basic routine, the score of his mathematics score will also get lower.

As future work, the aim is to expand the mapped skills, as well as the number of students and problems for the refinement of the results obtained. We also aim to compare the results obtained with those of other methods used to estimate skills. Finally, a practical application is to be implemented in a recommendation system, to observe the accuracy and effectiveness of the models in this environment.

We believe that a model that accurately evaluates skills and that allows mapping the evolution of sub-skills, incorporated into an educational recommendation system, offers the possibility of improving the indication of exercises appropriate to the needs of students, contributing positively to the processes of teaching and learning.

Acknowledgment

Universidade Federal do Rio Grande (FURG) and Instituto Federal de Educação, Ciência e Tecnologia Sul-riograndense (IFSul).

References

[1] F. B. Baker. *The basics of item response theory*. ERIC, 2001.
[2] S. C. Cazella, P. Behar, D. Schneider, K. K. da Silva, and R. Freitas. Desenvolvendo um sistema de recomendação de objetos de aprendizagem baseado em competências para a educação: relato de experiências. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 23, 2012.
[3] D. F. de Andrade, H. R. Tavares, and R. da Cunha Valle. Teoria da Resposta ao Item: conceitos e aplicações. *ABE, Sao Paulo*, 2000.
[4] R. J. De Ayala. *The theory and practice of item response theory*. Guilford Publications, 2013.
[5] A. E. Elo. *The rating of chessplayers, past and present*. Arco Pub., 1978.
[6] S. Embretson and S. Reise. *Item Response Theory*. Psychology Press, 2013.
[7] G. L. Moreira, W. Holanda, J. C. d. S. Coutinho, and F. S. Chagas. Desafios na aprendizagem de programação introdutória em cursos de TI da ufersa, campus Pau dos Ferros: um estudo exploratório. *Anais do Encontro de Computação do Oeste Potiguar ECOP/UFERSA (ISSN 2526-7574)*, (2), 2018.
[8] R. T. Nojosa. Teoria da Resposta ao Item (TRI): modelos multidimensionais. *Estudos em Avaliação Educacional*, (25):123–166, 2002.
[9] L. S. Oliveira. O desempenho em matemática do ENEM de 2012 em Luis Eduardo Magalhães (BA), na Teoria de Resposta ao Item. 2017.
[10] J. Y. Park, F. Cornillie, H. L. van der Maas, and W. Van Den Noortgate. A multidimensional IRT approach for dynamically monitoring ability growth in computerized practice environments. *Frontiers in psychology*, 10, 2019.
[11] L. Pasquali. *TRI–Teoria de Resposta ao Item: Teoria, procedimentos e aplicações*. Editora Appris, 2018.
[12] R. Pelánek. Applications of the elo rating system in adaptive educational systems. *Computers & Education*, 98:169–179, 2016.
[13] A. Prisco, R. Santos, S. Botelho, N. Tonin, and J. Bez. A multidimensional Elo model for matching learning objects. In *2018 IEEE Frontiers in Education Conference (FIE)*. 2018.
[14] A. Prisco, R. Santos, N. T. S. Botelho, and J. Bez. Using information technology for personalizing the computer science teaching. In *2017 IEEE Frontiers in Education Conference (FIE)*. 2017.
[15] C. C. Prodanov and E. C. de Freitas. *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico-2ª Edição*. Editora Feevale, 2013.
[16] M. D. Reckase. Multidimensional Item Response Theory. *Handbook of statistics*, 26:607–642, 2006.
[17] A. Robins. Learning edge momentum: A new account of outcomes in cs1. *Computer Science Education*, 20(1):37–71, 2010.
[18] C. M. Tavares. A teoria de resposta ao item na avaliação em larga escala: Um estudo sobre o exame nacional de acesso ao mestrado profissional em matemática em rede nacional-profmat. 2014.
[19] K. Wauters, P. Desmet, and W. Van Den Noortgate. Item difficulty estimation: An auspicious collaboration between data and judgment. *Computers & Education*, 58(4):1183–1193, 2012.