

Towards a Personalized Placement Assessment Method to Support Students' Knowledge Gaps Identification

Jerry Fernandes Medeiros^{1,2}, Marlon da Costa Monçores^{1,2}, and Samara Werner¹

¹ Tamboro Educacional, Rua Visconde de Caravelas, 111 - Botafogo, Rio de Janeiro

² Universidade Federal do Estado do Rio de Janeiro, Av. Pasteur, 458 - Botafogo, Rio de Janeiro

Abstract. Whereas students of basic education in Brazil advance in grade without mastering all the contents, teachers need to recall content from previous years. It can become irrecoverable as learners may have fallen behind their peers, or because middle or high school teachers may not have expertise in teaching basic academic skills. In this paper, we propose an adaptive assessment technique based on graphs for identifying learning gaps. The method described here does not rely on the calibration of items, but in the structure in which the content is taught, which may be customized according to the schools' pedagogical project.

Keywords: Adaptive Assessment, Computerized Adaptive Testing, Learning Gaps

1 Introduction

Educational assessment is the process of making reasonable inferences about what learners know based on evidence found from observation of what they say, do or make in selected situations [4].

The enhancement of the assessment process with adaptive features is important because it makes the assessment process more dynamic and individualized, as it adapts to the learner's performance. Furthermore, the number of questions required to estimate learners knowledge level is usually reduced, resulting in a less monotonous assessment process.

A Computer Adaptive Test (CAT) is a test administered by a computer where the presentation of each question, as well as the decision to finish the test are dynamically adapted based on the answers of the test takers. The test items are dynamically adjusted to a student's performance level. As a result, the test tends to be shorter and more accurate [7].

Numerous methodological approaches have been developed for CAT, most of them evolved from the principles of psychometric measurement. Linacre et al. [3] details the theoretical background involved in adaptive testing.

Numerous procedures have been developed for implementing the basic tasks needed to select and score adaptive tests. No method has been ideal for all assessing situations and scenarios. What works best depends on the unique characteristics of a given testing goal.

More directly, in adaptive testing methods, answering questions correctly or incorrectly implies in easier or more difficult questions being administered to the test taker. Questions are selected so that their difficulty matches the learner's inferred knowledge level in a given subject.

Questions that provide more information about what the learner knows are usually those with difficulty close to the learner's knowledge level. The learners knowledge level estimation depends on the number of questions answered correctly and on the difficulty level of the answered questions.

The technique approached in this paper uses a pool of questions which are highly structured and always related to a subject. It does not rely on the calibration of the items, but in the prerequisites defined by the school. The reasons for this decision will be addressed in the section 3.

An important consideration affecting the design of the assessment is the purpose on which it will be used. In this paper, we present a technique developed for finding learners knowledge gaps. The main purpose of this approach is to empower the teacher with the information needed for helping each learner to catch up with the desired progress.

2 Learning Gaps

A Learning gap is the existence of a difference between what a learner knows and what the learner is expected to know at a certain point in his education, as the age or scholar grade [1].

The learning gaps passed from one grade to another can become a significant problem that causes teachers to step in. Because of the gaps, teachers are usually forced to teach less on-grade content and focus on content from past grades. Closing these gaps can become an unsolvable long-term problem as the student learning gaps interfere with mastering new subjects.

Although some learning gaps are closed in following grades, the effort is high. As a consequence, less on-grade material can be discussed and learned.

One of the more consequential issues of the existence of learning gaps is their tendency to increase and become more severe over time. Furthermore, if basic skills such as reading, writing, and math are not acquired by students early on in their education, it may be more difficult for them to achieve it later on. As students progress through their education, closing learning gaps tend to become more difficult, because learners may have fallen behind their peers, or because middle or high school teachers may not have expertise in teaching basic academic skills. [1]

To bridge individual learning gaps, it is important that teachers have some tools that help them to identify and classify gaps in the classroom . In this context, the purpose of this paper is to describe an adaptive assessment method used

to identify learning gaps. The motivation and the context of use are described in section 3.

3 Application Scenario

The development of this work has focused on elementary school (Corresponding to the first nine years of basic education in Brazil).

The Programme for International Student Assessment (PISA) is a program that assesses random students aged 15, from around the world. It is applied every three years in public and private schools and was first implemented in 2000. The aim of PISA is to collect data in order to help in the development of policies to improve the education system of the countries that are enrolled.

The outcome of the maths test is divided into seven levels of proficiency ranging from "below level 1" to level 6. Each level is associated with a set of defined skills (except the "below level 1"). The seven levels and distribution of Brazilian students are shown in table 1

Level/Year	2003 accum. %	2006 accum. %	2009 accum. %	2012 accum. %
Bellow level 1	54.4	46.6	38.1	35.2
Level 1	21.7	25.9	31	31.9
Level 2	13.9	16.6	19	20.4
Level 3	6.5	7.1	8.1	8.9
Level 4	2.5	2.8	3.0	2.9
Level 5	0.8	0.8	0.7	0.7
Level 6	0.2	0.2	0.1	0.0

Table 1. The evolution of Brazilian students math skill over the years (PISA) [5] [6]

Based on this data, we can understand that almost all the students are on the first three levels of proficiency. By the cumulative percentages, it is clear that between 87.5% and 90% of the students are in the first three levels. This concentration in the lower region of the scale did not change significantly over the years, it may indicate the students are progressing over the grades but with some knowledge lacking. Brazilians students were ranked 58 out of 65 countries enrolled in the last edition.

The content administered to Brazilian students is in part guided by the Brazilian Law, which defines a common minimum curriculum that must be administered for all schools in the nation. In addition to this common curriculum, the school is responsible for a complementary curriculum. The aim of this complement is to address the needs and characteristics of the local community and it's culture [2].

It is expected that, when passing from one grade to the next, the student masters all content that was given in the previous year. However, for several reasons, this does not occur.

It is important to monitor student progress to prevent gaps becoming more pronounced with the passing years. This same law states that learner that have lost some content must catch up. However, it is up to the school, based on the principle of their autonomy and their right to define their pedagogical proposal, to decide how to address the bridging of students' gaps [2].

Given the presented scenario, where students progress to the next grade without mastering all required skills from previous years, and also considering the fact that there is not a unique curriculum, we have defined a method for assessing and finding students learning gaps. Our method does not rely on the calibration of the items, but on the structure of the content, allowing the possibility to easily customize the test to fit the schools' or teachers' needs.

4 Proposed Solution

This section describes the operation of the adaptive assessment for the gap discovery.

4.1 Graph Based Adaptive Assessment

We use a directed acyclic graph as the basis for the adaptive test. Let $G = (V, E)$ be a graph, where V is the set of vertices and E is the set of edges. Each vertex represents a subject that will be administered by the assessment and each edge represents a dependency relationship between two subjects. Thus, it is assumed that the source vertex is a prerequisite to the destination vertex. Cyclic dependencies are not allowed. It means that starting from a vertex V_1 there should not exist any other path that leads back to the vertex V_1 . This restriction is important because the existence of any cycle would make it impossible to define a precedence order among the elements belonging to the cycle.

The graph is constructed to represent the relationships between the subjects that are part of the content of a field of studies. Thus, the initial content of the subject are represented as initial graph vertices (no incoming edge) and are called roots. While subjects that are part of the final course are represented as end vertices (no outgoing edges) witch are called leafs. Other subjects are also represented as vertices, but they have incoming and outgoing edges, these are called only vertex.

In addition to the vertices and their relationships, one must associate the items (questions) to each of the vertices. Each vertex must have at least one item associated, however, there is no maximum limit. Figure 1 shows an example of a valid graph.

4.2 Setting up The Graph

After the graph creation with all subjects of the field and their relationships, it is recommended to pre-calibrate the graph. This pre-calibration can be made using different models of student profiles, i.e., students with low performance,

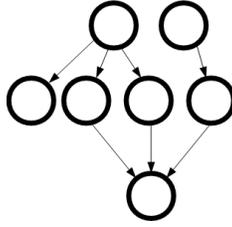


Fig. 1.

Example of a directed graph with 2 roots, 2 leaves, 3 other vertices, and no cycle

average performance and high-performance. Because of their different paces, these students also have different gaps. So, a good pre-calibration allows a more accurate gap identification with fewer items presented to the student during the assessment.

The pre-calibration algorithm aims to find out which vertices are more important to be asked for each student profile. In this regard, the algorithm simulates many students of a particular profile. For each student simulation, the algorithm simulates the student's answers according to his profile. The hit³ probability of an item in a vertex is inversely proportional to the amount of previous direct and indirect pre-requisites of the vertex, i.e., items that are in roots are more prone to be hit and the leaves are more prone to be missed. In addition, students with higher profiles are more likely to hit more questions.

The algorithm takes a graph G as a parameter. That graph contains the set of subjects and its relationship with each other. Another parameter is called *QUANTITY*, it is an integer and represents the number of students that will be used in the simulation. The third parameter is called *LEVEL* and it represents the average simulated student level and the last parameter is called *SD*, representing the standard deviation of the student's level.

After the simulation of each student, the algorithm then highlights the vertices that were in the student's knowledge borderline. Those vertices are identified by: (a) vertices that the student knows having edges to vertices that the student does not know or (b) vertices that the student does not know that are preceded by vertices that the student knows.

At the end of the simulation of all students, the algorithm calculates for each vertex, its *OPTIMAL VALUE (OV)* which is defined as a number of times the vertex appears in the border region, divided by the number of simulated students. Thus, the *OV* of a vertex is a number ranging from 0 to 1. When the *OV* is 0 it means the vertex does not belong to any border in any student and when the *OV* is 1 it means the vertex belongs to the border in all simulated students.

³ Every time a test taker answers correctly it is called a hit, on the other hand a wrong answer is called a miss

After pre-calibration, the *OV* for each vertex is stored and can be used for real students having their profile similar to the *LEVEL* used in the simulation.

4.3 Vertex Choosing

Vertex choosing is the central element of the adaptive assessment. A good vertex choice allows student's gaps to be found with few items. So, to make this choice more efficient, we used three elements combined.

1. ***OPTIMAL VALUE (OV)***: Represents the vertex percentage in the student's knowledge borderline. The higher *OV* the higher the vertex preference is in the selection process.
2. ***VERTEX CENTRALITY (VC)***: The betweenness centrality is calculated for each vertex. This value is a measure of centrality of a vertex in the graph. The higher *VC* the higher the vertex preference is in the selection process.
3. ***VERTEX SCORE (VS)***: Initially the score of all vertices is 0. Every answer given by the student lead to vertex score update. In general, the vertex score increases when the student hits a question and decreases when the student misses. Vertices with a value close to 0 are preferred in the selection process.

Our intention is to make each of those values equally weighted during the vertex choosing process. So, it is necessary that all elements have the same possible range of values. Thus, we decided to normalize its values between 0 and 1. For *OV* nothing needs to be done since their values are already in the defined range. However for *VC* it's necessary to normalize the value. This is done by dividing each vertex *VC* by the highest existing *VC*. So, the vertex with the higher *VC* will have the value of 1. The *VS* normalization is made in two steps: first, we pick up the absolute *VS*, then we make the operation $\frac{1}{1+VS}$. Thus, the values will always be between 0 and 1 and the priority goes to vertices with a score closer to 0.

The selection process also uses three constants that can be calibrated to increase or decrease the weight of any of those elements. The variables are called *OPTIMAL WEIGHT (OW)*, *Centrality WEIGHT (CW)* and *SCORE WEIGHT (SW)*. Thus, the value of each vertex is calculated according to the equation 1.

$$(OW * OV) + (CW * VC) + (SW * VS) \quad (1)$$

4.4 Graph Painting Technique

The goal of adaptive testing is to identify the student's learning gaps. For this identification, the algorithm proposed uses a technique that we have called graph painting. In this technique, each vertex has an associated value and a color. So, whenever a vertex has value 0, it also has the color black. When the value is

negative the vertex is red and when the value is positive the vertex color is blue. The semantics of each color is simple, blue vertices mean that the algorithm identified a student who knows the vertex subject. On the other hand, the red vertices are those where the algorithm identified a student who does not know the vertex subject. Black vertices are those where the algorithm has no information. Thus, the gap of the student is defined by all red and remaining black vertices at the end of the assessment.

When a student starts the assessment, all vertices are black and have it's VS equal to zero. Then, following the equation described in Section 4.3 a vertex is selected. The student then has to answer items related to the selected vertex. If there is more than one item associated with the vertex, the choice will be random, discarding items that the student has already answered previously (vertices can be revisited during assessment). Thus, the number of items displayed in each selected vertex is variable and depends on three factors: (1) number of available items in the vertex, (2) predefined maximum items than can be asked per vertex, and (3) student's performance. For the algorithm to consider that the student knows the vertex, it is necessary that the student hits all the questions that are shown to the vertex. So, if the student misses a question, the algorithm can change to another subject.

After the vertex completion, the algorithm assigns a value to the vertex. This value is controlled by the parameter *DIRECT WEIGHT (DW)*. The assigned value may be DW if the algorithm considers that the student knows the subject or $-DW$ otherwise. This change in vertex value will cause the vertex to be painted as blue or red. In addition, in painting the selected vertex, the algorithm also changes the value of some other vertices following a simple rule. If the selected vertex is painted blue, all predecessors vertices are also painted blue. If the selected vertex is painted red, all successors vertices are also painted red. However, the value assigned to these other vertices is defined by the parameter *INDIRECT WEIGHT (IW)* which, by default, is set as $DW * 0.1$. Figure 2 shows two examples of graph painting algorithm. The first example shows the result after a hit and the second shows a result after a miss. The green dot at the center represents the selected vertex in each situation.

4.5 Graph Based Model Benefits

As the subjects are mapped in a directed graph, the algorithm can infer information about subjects that weren't asked, creating a possibility to evaluate the gap of a broad curriculum asking only a few items to students. Another benefit is that the items do not need to be calibrated because the most important factor is the relationships between subjects. Thus, it is possible that organizations using a different subject teaching order use the same set of items. The only adjustment needed is to configure the specific graph, that is, the graph can be set to identify gaps in the same order that they are presented in a specific course.

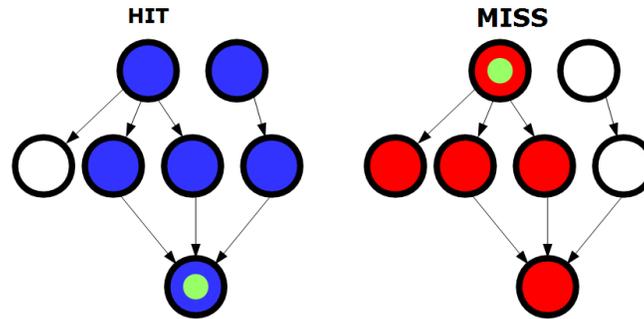


Fig. 2.

Example of a graph being painted after a hit and after a miss. The green center shows the selected vertex.

4.6 Considerations and Future Work

In final considerations, we emphasize that the research is at an early stage. This proposed test model has not yet been applied to real students. However, we already have a graph containing more than 5000s item from all maths subjects taught in years 5-9 in Brazilian elementary school. Thus, the next step in the research is applying an assessment test using the described algorithm to students who are enrolled in these scholar grades.

In our preliminary study we are simulating students with different paces and performances from different grades. The algorithm is able to paint more than 80% of the graph with less than 30 subjects selections.

We are also working on reports, based on the results obtained from the assessment. These reports will help teachers in decision-making about tutoring classes for addressing the gap .

References

1. Abbott, S., Guisbond, L., Levy, J., Sommerfeld, M.: The glossary of education reform. Hidden curriculum. Retrieved (2014), <http://edglossary.org/learning-gap/>
2. de Diretrizes, L.: bases da educação nacional (1996)
3. Linacre, J.M., et al.: Computer-adaptive testing: A methodology whose time has come. Chae, S.-Kang, U.-Jeon, E.-Linacre, JM (eds.): Development of Computerised Middle School Achievement Tests, MESA Research Memorandum (69) (2000)
4. Pellegrino, J.W., Chudowsky, N., Glaser, R., et al.: Knowing what students know: The science and design of educational assessment. National Academies Press (2001)
5. de Estudos e Pesquisas Educacionais Ansio Teixeira (Inep), I.N.: Resultados nacionais pisa 2006 (2008)
6. de Estudos e Pesquisas Educacionais Ansio Teixeira (Inep), I.N.: Relatrio nacional pisa 2012 (2012)
7. Thissen, D., Mislevy, R.J.: Testing algorithms. Computerized adaptive testing: A primer 2, 101–133 (2000)