2021

The Use of Rasch Model to Create Adaptive Practices in e-Learning Systems

Wisam Zaqoot
National University of Singapore, wisamzaqoot@comp.nus.edu.sg

Lih-Bin Oh
National University of Singapore, ohlb@comp.nus.edu.sg

Elizabeth Koh
National Institute of Education, Nanyang Technological University, elizabeth.koh@nie.edu.sg

Lay Hoon Seah
National Institute of Education, Nanyang Technological University, layhoon.seah@nie.edu.sg

Hock-Hai Teo
National University of Singapore, teohh@comp.nus.edu.sg

The Use of Rasch Model to Create Adaptive Practices in e-Learning Systems

Full research paper

Wisam Zaqoot
Department of Information Systems and Analytics
National University of Singapore
Singapore, Singapore
wisamzaqoot@comp.nus.edu.sg

Lih-Bin Oh
Department of Information Systems and Analytics
National University of Singapore
Singapore, Singapore
ohlb@comp.nus.edu.sg

Elizabeth Koh
National Institute of Education
Nanyang Technological University
Singapore, Singapore
elizabeth.koh@nie.edu.sg

Lay Hoon Seah
National Institute of Education
Nanyang Technological University
Singapore, Singapore
layhoon.seah@nie.edu.sg

Hock-Hai Teo
Department of Information Systems and Analytics
National University of Singapore
Singapore, Singapore
teohh@comp.nus.edu.sg

Abstract

While different approaches were developed to create computerized adaptive practices for e-learning systems, we show that exploiting Rasch models to create adaptive practices can be a new promising approach. Rasch analysis enables us to find a mathematical model to analyze students' answers to exam questions by representing students' abilities and questions difficulty levels on the same scale. In this paper, we introduce a novel algorithm to generate adaptive practices based on the Rasch analysis of students' performance in an initial assessment. This approach enables us to generate adaptive practices that consider not only the student's ability and his previous performance but also the difficulty level of each question. We also present results from a preliminary field experiment that we have conducted using an online learning system that implements this algorithm. The potential advantages of this approach and the practical contributions are discussed.

Keywords e-learning, Rasch model, adaptive practices, personalized education

1 Introduction

Terms like adaptive learning, adaptive teaching, and adaptive education gained prominence among educators and researchers. A lot of effort was exerted on developing adaptive learning materials, adaptive practices, adaptive feedback, and adaptive testing techniques. Researchers, for example, tried to personalize learning materials or tests difficulty according to learners' personality, preferences, style of learning, prior performance, or prior knowledge (Landsberg et al. 2012; Nakic et al. 2015; Xie et al. 2019). Some studies also tried to assess adaptive learning and showed that it could be beneficial for learners (Liu et al. 2017; Nakic et al. 2015; Yang et al. 2013).

In this paper, we focus on adaptive e-learning practices, where quizzes and training practices are generated according to each student's learning needs. In fact, there are many techniques that can be used to pick a set of questions from a question bank in order to create such adaptive practices. However, we found that these techniques are either too simple as they give the learners questions similar to the ones they failed earlier or very complicated and computationally expensive as they depend on sophisticated statistical or machine learning techniques that need to be integrated into the learning systems. Furthermore, a common weakness in these approaches is that they try to customize the practice quizzes taking into consideration the learners' performance while neglecting the difficulty of the questions. These techniques may also ignore the broader picture of the whole class performance when personalizing practices for a specific student. Such adaptive practices would easily be misled if the student gave a correct answer by chance for a difficult question or a wrong answer by mistake to an easy question.

Hence, we propose a new approach for selecting questions adaptively from a question bank to create adaptive practices. Our proposed algorithm is based on Rasch analysis, which enables us to identify and contrast overall students' abilities and difficulty levels of the different questions before creating a personalized practice for a specific student. Henceforth, we anticipate that our algorithm would be able to overcome the shortcomings of the existing techniques. In fact, Rasch analysis is a powerful tool that is popular in education, psychometric and business research where it is used to construct and validate instruments and to compute respondents' performances (Boone 2016). When it comes to constructing instruments, Wulf and Winkler (2020) showed that Rasch analysis can overcome some of the shortcomings of the confirmatory factor analysis (CFA). Unfortunately, Rasch analysis has not been widely adopted in the IS literature yet.

2 Rasch Analysis in Education

Rasch analysis enables us to find a mathematical model for a latent trait while accomplishing probabilistic conjoint additivity. It is 'conjoint' as it measures persons and items on the same scale, and it is 'additive' as the scale has equal intervals (Bergner 2017; Granger 2008). We can use the Rasch analysis on raw scores (from a survey, scale or test, etc.) to compute linear "person measures" that express the performance of the respondents, and to compute the difficulties across all test or survey items (Boone 2016). This means that Rasch analysis can be used in education to analyze scores in an exam that covers some competency while trading off students' abilities against the difficulty of questions in that exam. From this standpoint, a Rasch model will present the ability level of each student expressed as a number along an infinite linear scale of the relevant ability. Similarly, each question is characterized by a difficulty level also expressed as a number along the same infinite scale of the relevant ability (Linacre 2000). If we have n students, the ability of student i is identified as B_i units from the local origin on that scale. Likewise, if we have m types of questions, the difficulty level of question type j is identified as D_j. The relationship between the student ability and the question difficulty can be expressed by the (dichotomous) Rasch model using the following Equation 1 (Dekleva and Drehmer 1997):

$$ln\left(\frac{P_{ij}}{1-P_{ij}}\right) = B_i - D_j \quad \text{----------------------------} \ (Eq.1)$$

Where P_{ij} is the probability of student i answering the question type j correctly, and $1 - P_{ij}$ is the probability of answering it incorrectly. From the previous equation, we can also derive P_{ij} as follows:

$$P_{ij} = \frac{e^{(B_i-D_j)}}{1+e^{(B_i-D_j)}} \quad \text{------------------------------} \ (Eq.2)$$

For example, if a question happened to be placed along the ability metric of a specific student (i.e. $B_i = D_j$), from the previous equation, we can see that this student is expected to have a 50% chance to answer this question correctly. And if the student's ability, for example, is one logit above the question difficulty (i.e. $B_i - D_j = 1$), then the probability of the student answering this question correctly is 73.1%.

2.1 Rasch Model and Wright Maps

While the measures of the ability and difficulty are given by the Rasch model as exact numerical logit values, it can be easier to visualize these measures using a Wright map like the one shown in Figure 1. In such Wright maps, students' abilities and question difficulty are represented along the same scale of the competency we are measuring. The more able students and the more difficult questions are placed near the top of the map, while the less able students and the easier questions are placed near the bottom. It is also common to show measures as logits on the map and to show the mean level, and one or more standard deviations above and below the mean.

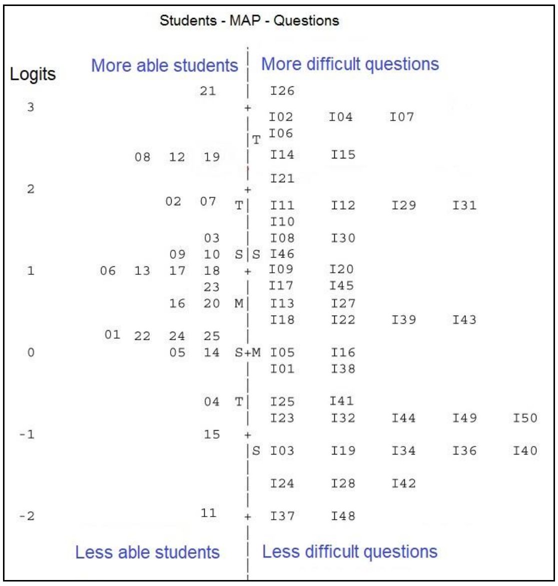

Figure 1. Wright Map Example

2.2 Rasch Analysis and Computer Adaptive Tests

Besides evaluating students' abilities and questions difficulty in an exam, another common use for Rasch analysis is to create Computer Adaptive Tests (CATs). The general idea behind CATs is to measure the ability of the test-taker quickly by selecting questions (from a question bank) that match the estimated ability of the test-taker. If the test-taker succeeds on a question, a slightly more challenging question is presented next, and vice-versa. The process repeats, and after each answer, the ability estimate of the test-taker is revised (see Figure 2). This algorithm should quickly converge into a sequence of questions matching the test-taker's effective ability level (Linacre 2000). The selection of questions adaptively from the question bank is based on the Rasch model, which estimates the difficulty levels of the questions against the test-taker's ability.

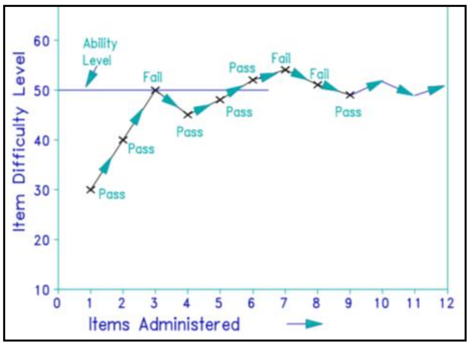

Figure 2. Dichotomous CAT Test Administration (Linacre 2000)

Over the last few decades, several CAT algorithms were developed according to this general approach. However, these algorithms may differ in their details like the initial question selection, content balancing, amount of estimate revision after each answer, test length, and the test stopping rules.

3 Computer Adaptive Practices

In contrast to the popularity of adaptive testing based on Rasch modelling and Item Response Theory (IRT) in general, we can hardly find any reference to the use of Rasch models to create adaptive practices. However, we believe that Rasch analyses are very promising in this area and have a great potential to be used to develop adaptive quizzes, practice tests, and training materials.

Similar to CATs, we consider the case of having learning systems using Computer Adaptive Practices (CAPs) to give learners customized questions from a question bank while taking learners' ability into consideration. The main difference between CATs and CAPs is in the goal and how to achieve it. While CATs aim to measure the learner's ability by determining questions with a level of difficulty matching the learner's ability, the CAPs aim to enhance the learner's ability by giving the learner training questions that are around and above his estimated ability.

Chrysafiadi & Virvou (2013) identified nine types of student modeling techniques that are used to create adaptive learning systems, including overlay model, perturbation student model, Constraint-Based Model (CBM), Bayesian networks and machine learning techniques. In their exhaustive survey, the authors classified sources of adaptation into: knowledge, errors and misconceptions, learning styles and preferences, cognitive aspects, affective features, motivation and meta-cognitive characteristics (Nakic et al. 2015). Surprisingly, we found that Rasch analyses, despite their great potential, received little attention when it comes to creating adaptive content or training materials on digital learning systems, with very few exceptions like Brinkhuis & Maris (2020) and Klinkenberg, Straatemeier, & Van Der Maas (2011).

3.1 Proposed Algorithm

Considering this research gap and the great potential for using Rasch analyses in this field, we try in this paper to develop a new approach for CAPs that is based on Rasch analyses. According to the algorithm we propose, scores from a pre-test can be analyzed using Rasch analysis to create a Rasch model. Hereafter, we can use this model to identify the question types that are of difficulty level just around or above the ability of each student, as these are the question types that we want the student to practice. Figure 3 shows the proposed algorithm.

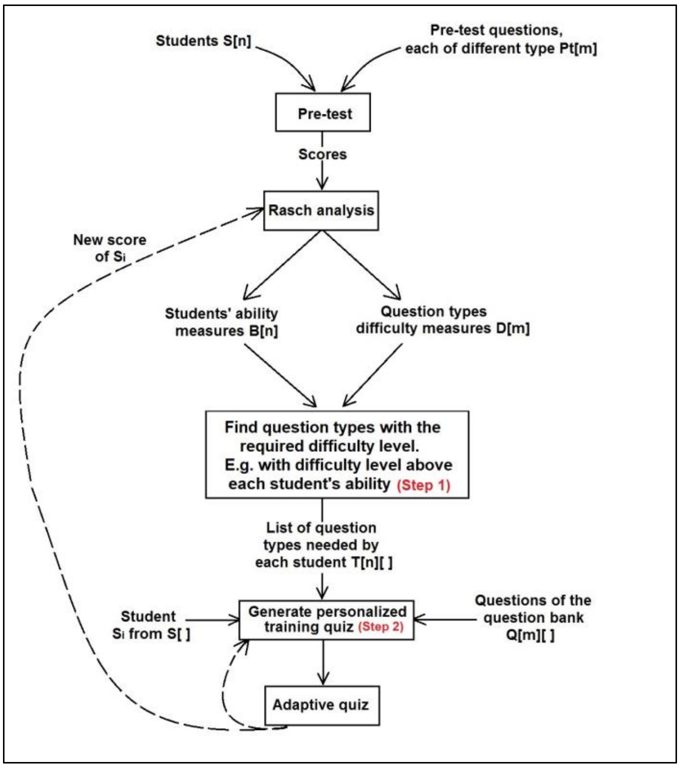

Figure 3. The Proposed Algorithm

This algorithm is based on having a question bank consisting of different questions that belong to a specific set of question types. While questions of different types may differ in their difficulty, questions under each type should have a high degree of similarity and hence are assumed to have the same level of difficulty.

One important step in this algorithm (denoted as step 1 in the graph above) is meant to adaptively find the question types that should be given to each student after considering students' abilities and the difficulty level for each question type. In the most basic approach, the algorithm would select only question types with difficulty levels above the ability of the student, as detailed in the following pseudocode and shown in Figure 4(a):

```
For each B_i in B[n]
    For each D_j in D[m]
        If D_j > B_i Then
            Add t_j to T[i][ ]
        End If
    Endfor
Endfor
```

Where t_j is a question type that is one of the m question types we have, and T is the list of question types the students need to practice.

Alternatively, we may customize the algorithm by changing the threshold we want for picking questions. For example, we may choose a threshold along each student's level in order to pick question types with difficulty at least equal to student's ability (as shown in Figure 4 (b)), or even a threshold that is below student's ability to some degree (as shown in Figure 4 (c)). In this case, if we choose a threshold, for example, at 1.4 logits below each student's ability, the algorithm will select question types of difficulty at least 1.4 logits below student's ability, which represent questions that the student has a probability of no more than 80% of answering them correctly (refer to Equation 2). Similarly, if we want to train students on all question types that they have less than 90% chance of answering them correctly, the algorithm should have a threshold at 2.2 logits below each student's ability.

An alternative approach would give students questions of all types, however, the question types above student's ability would be given a far higher probability of appearing in the practice quizzes in comparison to questions below student's ability. In this case, the student may see training questions of all types, but the majority of questions will be of the difficult types.

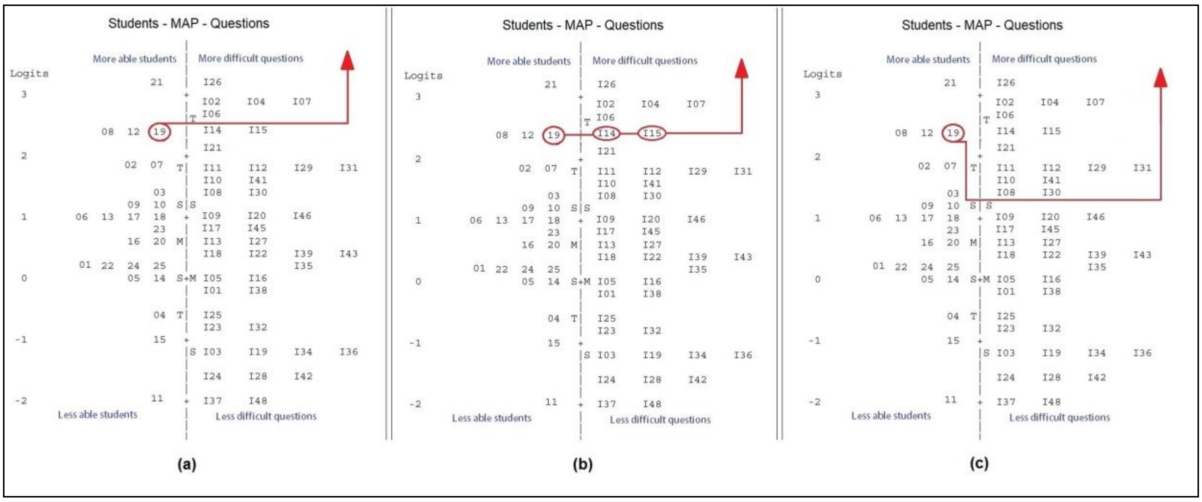

Figure 4. Examples of Different Thresholds for Selecting Question Types for Student #19

Another stage where our general algorithm can be customized is denoted as step 2 in Figure 3. In this step, the generation of the personalized quiz can be accomplished in many different ways. For example, we may generate quizzes of a different number of questions according to the number of questions the student needs to practice. Alternatively, we may generate quizzes of a fixed number of questions. In this case, we need to decide how many times we want to repeat questions of the same type. We may also need to decide what to do if the question types selected for a student are exhausted, should we repeat the questions already taken, or consider giving students random questions from the question bank.

After taking the adaptive quiz, we have two possible approaches (denoted in the dotted arrows in Figure 3). One approach is to return to the previous step to regenerate a new adaptive quiz using the same question types we used earlier for this student. Another approach is to take the student's new scores from this quiz to update the Rasch model on the fly so that we can generate a new adaptive quiz with a new set of question types matching the student's updated ability.

4 Research Method

4.1 Experiment Design

As a part of a multi-year project, we developed a test to measure the representational fluency of elementary school students. Representational fluency is defined as the ability to reason and work among multiple representations. The representations we focus on include visual representations like tables and various graphs and diagrams found in school science curriculum. Our test consists of 25 different types of questions; all of them are multiple-choice questions (Zaqoot et al. 2019a, 2019b).

We also developed an online learning system to teach students the skills they need to be fluent in dealing with these representations. One part of the learning system we developed is the practice quizzes part. We prepared a question bank that includes about 300 questions. Quizzes generated from this question bank are given to students to practice the representational fluency skills.

We conducted multiple experiments using the learning system we developed. The experiments were conducted on grade 5 students in an elementary school. Since we are conducting multiple experiments at the same time, we splitted 230 students into multiple groups. For the purpose of this experiment, we created a treatment group by randomly selecting 57 students from the school. We also randomly selected a control group of 59 students.

Complying with the research ethics, we gave students and their guardians the choice not to participate in the experiment and the ability to opt out at any point. As a result, we began the experiment with 37 students in the treatment group and 44 students in the control group. Unfortunately, 6 more students in the control group dropped out before completing the experiment. Eventually, our sample consisted of 37 students in the treatment group and 38 students in the control group completing the experiment.

4.2 Experiment Procedures

Students in the treatment and control groups began by taking a pre-test to measure their representational fluency using our online learning system. For the following two months, they used the learning system to take lessons about the representational fluency skills followed by practice quizzes. Each practice quiz consisted of 12 questions covering different representational fluency skills. Each student has to take at least two practice quizzes.

Students in the treatment group were given adaptive practice quizzes generated using a customized version of the algorithm we presented earlier in this paper. The details of this customized algorithm are presented in the following subsection. On the other hand, students in the control group were given practice quizzes with questions chosen randomly from the question bank. Hence, the performance of students in the control group represents a baseline that can be contrasted to the performance of those in the treatment group taking adaptive practices.

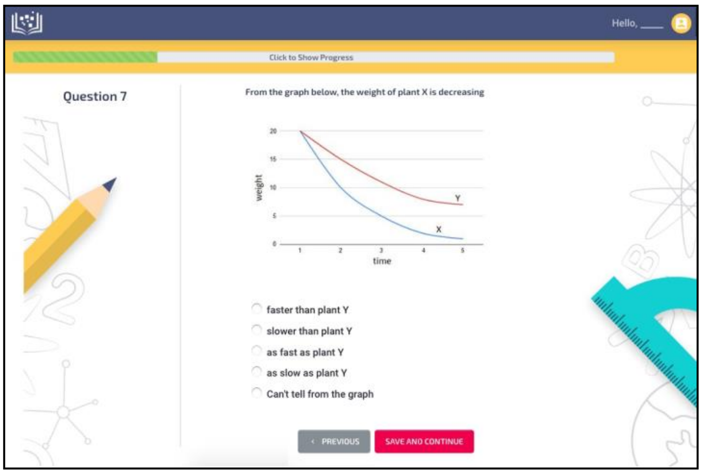

Figure 5. An Example Question from the Practice Quizzes in the Learning System

By the end of the training program, students were given another representational fluency test as a post-test to measure their performance against their scores in the pre-test. Table 1 summarizes the experiment procedures:

Control Group	Treatment Group
Pre-test	Pre-test
10 lessons learning material	10 lessons learning material
Practices questions selected randomly by the system	Adaptive practices questions selected by the adaptive algorithm
Post-test	Post-test

Table 1. Experiment Design

4.3 Customized Algorithm and Quizzes Generation

While the algorithm we used in this experiment follows the general design we presented earlier in this paper, we made several design choices while developing our learning system. For example, we decided to conduct the Rasch analysis using the students' scores in the pre-test. In order not to complicate our experiment design, we conducted the Rasch analysis manually once and did not update the Rasch model on the fly after each quiz was taken.

In addition, since the probability of answering a question correctly by a student is only 50% if this student is placed along with the difficulty metric of that question, we decided to choose the algorithm threshold for picking questions to be slightly below student's ability in order to pick question types of a difficulty level that is just below student's ability or higher (as shown in Figure 4 (c))

Finally, we decided to give students the choice to take as many practice quizzes as they wish, with a minimum number of two quizzes. The learning system will use the list of training question types suggested by the algorithm to generate practice quizzes of 12 questions each for every student in the treatment group. It is allowed for different questions of the same type to appear in the same quiz.

4.4 Results and Data Analysis

Our main hypothesis here is that giving students adaptive practices generated using our algorithm that is based on Rasch modeling will enable students to enhance their representational fluency skills. This can be reflected in students achieving higher scores in the post-test. Hence, we hypothesize that:

Students in the treatment group (taking adaptive quizzes) will perform better than students in the control group.

The summary of experiment results is detailed in Table 2:

	Control group	**Treatment group**
N (gender)	38 (18 F + 20 M)	37 (16 F + 21 M)
Post-test scores mean	16.84	17.81
Post-test scores SD	4.11	3.64
Minimum post-test score	6	7
Maximum post-test score	23	24
Average number of quizzes taken	2.02	2.68
Average duration of post-test	552.3 seconds	625.7 seconds

Table 2. Experiment Descriptive Statistics

Preliminary results show no significant difference between the two groups in students' post-test scores as $t(73)=1.079, p=0.284$.

Since we are conducting a two-group pretest-posttest experiment design, we decided to use analysis of covariance (ANCOVA) to analyse the outcomes. However, ANCOVA has some assumptions (Glass et al. 1972) and before conducting the ANCOVA analysis we need to ensure that the assumptions of this analysis are met.

ANCOVA main assumptions are:

1. Independence of observations
2. Homogeneity of variance
3. Homogeneity of regression slopes
4. Linearity (of the relation between the covariates and the dependent variable)

For the homogeneity of variance assumption, Levene's Test indicated equal variances *(F=.017, p=.897)*. There was also no significant difference between the treatment and control groups on the pre-test scores *(F=1.463, p=.230)*. Moreover, for the homogeneity of regression condition, we found no significant interaction between the groups and the pre-test *(F=.373, p=.543)*. We can also see in Figure 6 that the homogeneity of regression slopes and linearity assumptions seem to be met. Meeting these assumptions means that we can conduct the ANCOVA analysis.

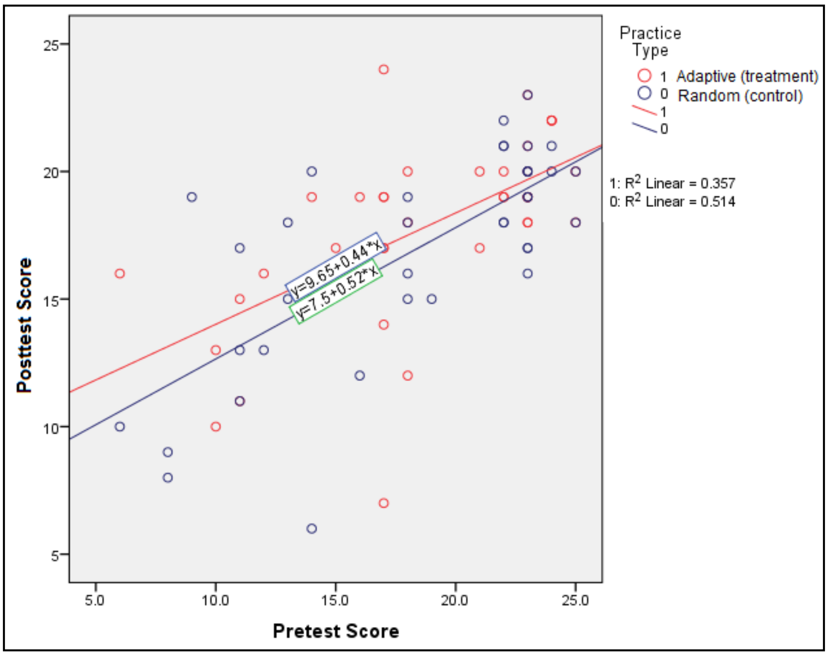

Figure 6. Regression lines for treatment and control groups

Applying the ANCOVA to evaluate the difference in post-test scores between the treatment and control groups, after controlling for the pre-test scores, revealed no significant main effect for the treatment since $F(1,72)=1.055$, $p=.308$, $\eta_p^2=.014$.

We also tried to conduct ANCOVA again after controlling for other variables like gender, the number of practice quizzes the student took, the time spent taking the lessons, the duration of quizzes, etc. However, results continued to show no significant effect for the intervention.

Additionally, we decided to conduct a difference in difference (DID) analysis for the results to evaluate the effect of the treatment. DID can help us in cases where there are inherent unobserved differences between the two groups that we did not overcome by randomizing treatment assignment over such small samples. The DID regression we used is simply:

$$PosttestScore_{it} = \beta_o + \beta_1 Post_t + \beta_2 Adaptive_i + \beta_3 Post_t * Adaptive_i + \varepsilon_{it} \quad \text{--------------------} \quad (Eq.3)$$

Where $Post_t$ is a dummy variable that equals 1 for the post-test, $Adaptive_i$ is a dummy variable that equals 1 for being in the treatment group, $Post_t * Adaptive_i$ is the interaction between the two variables and β_3 is the DID estimate we want to find out.

By regressing the variables, we found that β_3 was *-.050, p=.717*. This coefficient remained insignificant even after we added some control variables (like number and duration of practice quizzes taken). Hence, we are not able to find a statistically significant effect for the treatment.

Overall, these results mean that our main hypothesis is rejected. Looking at the experiment outcomes analyses, we find that utilizing adaptive practices in our training program resulted in a minor enhancement in students' performance in the post-test that was not statistically significant. However, failing to find a statistically significant result does not suggest that our treatment was not effective, but rather it may indicate that the data we collected did not include a sufficient amount of evidence to prove the effectiveness of our treatment (if it is indeed effective) (Mertens and Recker 2020). While it is possible that our treatment was not effective, we still believe in the value of our algorithm and see that it has the potential to help learners by giving them truly adaptive practices that can help them. Hence, we decided to review the experiment design, experiment handling, and the outcomes again for possible reasons for the insignificant outcome. In fact, we found some sources of concern and confounding factors that should be considered in future experiments to assess our algorithm more carefully. For example, we found that:

1- One concern we have is related to the sample size. Many students opted out before and during the experiment. We ended up with 37 students in the treatment group and 38 students in the control group. The average score in the post-test was 17.81 for the treatment group while it was 16.84 for the control group. This small improvement was less than what we were expecting, and it turned to be too small to be significant statistically. It is likely that we will find a significant difference between the two groups if we had larger samples.

2- Our experiment design included giving students in both groups the same 10 lessons covering the skills we want them to master, before letting them take the practices. However, if the lessons were good enough, their effect may overshadow the effect of taking practice quizzes. In other words, if the students in both groups learned almost everything they need from the lessons, giving students adaptive or random quizzes afterwards would hardly make any difference between the two groups.

3- It is possible that the topic we tried to teach to students in this learning program was not appropriate. We tried here to give students the competences they need to be '*fluent*' in dealing with different visual representations they have in their curriculum. Becoming fluent in these skills may require intensive training that exceeds what we offered in our learning program.

4- While we attempted to create learning materials and practice quizzes that help students master the representational fluency skills examined by the pre- and post-tests, there is still a possibility that our post-test was not sensitive enough to capture the improvement in students' skills after the training they had. In this sense, the insignificant outcomes we found could be the result of test insensitivity rather than treatment ineffectiveness.

5- While the assignment of treatments in our experiment was randomized, there is still a potential source of self-selection bias that stems from the research ethics requirement to allow students and their parents to decide whether the student would participate in the experiment and allowing students to withdraw from the experiment at any time. For example, a follow-up analysis showed that the six students who withdrew from the experiment after completing the pre-test (all of them from the control group) were performing far lower than their peers (their average pre-test score was only 11.0). Their withdrawal raised the average performance for the control group. It is also reasonable to assume that the other 35 students who decided not to participate before the experiment commencement may differ significantly in their abilities from those who participated.

In addition, there is another interesting possibility that we may want to consider, that the effectiveness of our adaptive practices is not necessarily reflected in an improvement in students' performance in the post-test, but may be reflected in other aspects. For example, there are some indicators of a higher commitment and/or enjoyment among students in the treatment group in comparison to those in the control group. For instance, no students in the treatment group withdrew after the experiment started (in comparison to 6 students in the control group). Furthermore, students in the treatment group took on average 2.68 practice quizzes (2.02 only for those in the control group). We also found that students in the treatment group spent 445 seconds on average per practice quiz (361 seconds in the control group). Albeit this can also be due to the fact that the control group practice quizzes tend to be easier as their questions were selected randomly and may include very easy questions. Hence, we believe there is a need for additional experiments that take the previous concerns into consideration, in order to affirm the internal and external validity of our approach. We also believe that performing further learning analytics would enable us to derive more interesting insights.

5 Conclusion

Several studies emphasized the benefits of adaptive learning in general. Researchers also worked on developing different approaches to create adaptive learning materials, practices and tests. Adaptive practices, for example, try to enhance the learners' experience by giving them practice materials that fit their learning needs while saving them time and effort by focusing only on the learners' weaknesses and the skills that they need to master while avoiding irrelevant questions.

Even though there are several approaches for selecting questions from a question bank to create CAPs, we decided to develop a new approach to generate adaptive practices. This approach is based on the Rasch analysis of a pre-test, where we use the Rasch modeling of pre-test scores to determine question types that have a difficulty level around or above the learner's ability. We argue that this novel approach can be superior to the current approaches in many contexts for the following reasons:

1. By using Rasch modeling, we are less tied to student's answers in the pre-test, and more able to create adaptive practices that take student's ability and the overall difficulty of questions into consideration. This can be an important feature, considering that students sometimes give right answers by guessing or wrong answers because of carelessness. The Rasch analysis can mitigate this problem by looking into the overall performance of the class and the difficulty levels of the questions, and hence, partially overlooking whether a specific student answered a question right or wrong.
2. In contrast to other approaches used to create CAPs that demand a high computational cost like machine learning and Bayesian networks, we expect the use of Rasch analysis to be less computationally intensive.
3. One main shortcoming in some current approaches, especially those depending on machine learning, is that they require a large training dataset. Such large datasets may not be available in many learning contexts. In contrast, effective Rasch models can be created even for very small samples of learners.
4. In contrast to many other approaches, it can be much easier for the teacher to understand why a CAP based on a Rasch model suggested a specific set of questions for a specific student. Here, a Wright map can be used to explain the outcomes of the adaptive algorithm.

On the other hand, we acknowledge that the use of Rasch analysis in this field does not come without problems. For example, one main limitation with Rasch analyses is that they assume the unidimensionality of the test. This strict assumption can be problematic since most tests are multidimensional. However, we can tolerate this assumption to some degree as long we can achieve approximation to unidimensionality (Linacre 2000).

We hope that our novel approach to create CAPs using the new algorithm we introduced in this paper will give a new powerful technique for practitioners in the field of developing adaptive e-learning systems and will help learners by giving them better CAPs that are more effective in addressing their learning needs. However, the preliminary experiment we administered to examine the effectiveness of CAPs generated using our algorithm yielded modest outcomes that were statistically insignificant. A deeper examination of the experiment results raised many issues and questions that need to be considered in future experiments in order to assess the effectiveness of our approach more rigorously.

Nonetheless, we still believe that the new approach of using Rasch models to create CAPs can open a new direction for researchers eager to develop more intelligent adaptive learning systems. Furthermore, the general design of the algorithm presented in this paper leaves plenty of room for future improvements, customization and tuning to create more efficient and effective versions of this algorithm.

6 References

Bergner, Y. 2017. "Measurement and Its Uses in Learning Analytics," in *Handbook of Learning Analytics*, pp. 35–48.

Boone, W. J. 2016. "Rasch Analysis for Instrument Development: Why, When, and How?," *CBE-Life Sciences Education* (15:4), American Society for Cell Biology.

Brinkhuis, M. J. S., and Maris, G. 2020. "Dynamic Estimation in the Extended Marginal Rasch Model with an Application to Mathematical Computer-adaptive Practice," *British Journal of Mathematical and Statistical Psychology* (73:1), pp. 72–87.

Chrysafiadi, K., and Virvou, M. 2013. "Student Modeling Approaches : A Literature Review for the Last Decade," *Expert Systems With Applications* (40), Elsevier Ltd, pp. 4715–4729.

Dekleva, S., and Drehmer, D. 1997. "Measuring Software Engineering Evolution: A Rasch Calibration," *Information Systems Research* (8:1), pp. 95–104.

Glass, G. V, Peckham, P. D., and Sanders, J. R. 1972. "Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance," *Review of Educational Research* (42:3), pp. 237–288.

Granger, C. V. 2008. "Rasch Analysis Is Important to Understand and Use for Measurement," *Rasch Measurement Transactions*, pp. 1122–1123. (https://www.rasch.org/rmt/rmt213d.htm, accessed May 20, 2021).

Klinkenberg, S., Straatemeier, M., and Van Der Maas, H. L. J. 2011. "Computer Adaptive Practice of Maths Ability Using a New Item Response Model for on the Fly Ability and Difficulty Estimation," *Computers & Education* (57:2), pp. 1813–1824.

Landsberg, C. R., Astwood Jr, R. S., Buskirk, W. L. Van, Townsend, L. N., Steinhauser, N. B., and Mercado, A. D. 2012. "Review of Adaptive Training System Techniques," *Military Psychology* (24:2).

Linacre, J. M. 2000. "Computer-Adaptive Testing: A Methodology Whose Time Has Come," *MESA Memorandum* (69).

Liu, M., McKelroy, E., Corliss, S. B., and Carrigan, J. 2017. "Investigating the Effect of an Adaptive Learning Intervention on Students' Learning," *Educational Technology Research and Development* (65:6), pp. 1605–1625.

Mertens, W., and Recker, J. 2020. "New Guidelines for Null Hypothesis Significance Testing in Hypothetico-Deductive IS Research," *Journal of the Association for Information Systems* (21:4), pp. 1072–1102.

Nakic, J., Granic, A., and Glavinic, V. 2015. "Anatomy of Student Models in Adaptive Learning Systems: A Systematic Literature Review of Individual Differences from 2001 to 2013," *Journal of Educational Computing Research* (Vol. 51).

Wulf, J., and Winkler, T. J. 2020. "Evolutional and Transformational Configuration Strategies: A Rasch Analysis of It Providers' Service Management Capability," *Journal of the Association for Information Systems* (21:3), pp. 574–606.

Xie, H., Chu, H., Hwang, G., and Wang, C. 2019. "Trends and Development in Technology-Enhanced Adaptive/Personalized Learning : A Systematic Review of Journal Publications from 2007 to 2017," *Computers & Education* (140), Elsevier.

Yang, T., Hwang, G., and Yang, S. J. 2013. "Development of an Adaptive Learning System with Multiple Perspectives Based on Students' Learning Styles and Cognitive Styles," *Journal of Educational Technology & Society* (16:4), pp. 185–200.

Zaqoot, W., Oh, L.-B., Seah, L. H., Koh, E., Zhou, F., Tan, W., and Teo, H.-H. 2019a. "Representational Fluency in Education: A Literature Review and the Proposal of a New Instrument," in *IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE2019)*, Yogyakarta.

Zaqoot, W., Oh, L.-B., Seah, L. H., Koh, E., Zhou, F., Tan, W., and Teo, H.-H. 2019b. "Development and Validation of a Representational Fluency Test for Primary School Students," in *The Tenth International Conference on Technology for Education (T4E)*, Goa.

Acknowledgements

The authors acknowledge the financial support provided by the Singapore Ministry of Education Academic Research Fund and the valuable help and support from the participating school.

Copyright