Personalization to Student Interests in Reasoning Mind: Depth, Grain Size, and Ownership

Candace Walkington
Southern Methodist University
3101 University Blvd., Ste. 345
Dallas, TX, 75205
1-512-417-9975
cwalkington@smu.edu

Leigh A. Mingle
Reasoning Mind
2000 Bering Drive #300
Houston, TX 77057
1-281-579-1110
Leigh.Mingle@reasoningmind.org

ABSTRACT

In this paper, we describe a study conducted within the Reasoning Mind blended learning mathematics curriculum for 6th grade. This study explores the instructional principle of *context personalization*, where learning materials are connected to students' out-of-school interests in topics like sports and video games. In the study, students received either (1) normal story problems, (2) problems that were personalized to their out-of-school interests assigned to them by the computer based on prior interest survey ratings, (3) personalized problems that were randomly assigned to students by the computer regardless of their rated interests, and (4) personalized problems where the student could choose the topic of the problem. We present analyses of how the different conditions performed in terms of their immediate performance and long-term learning, interpreting results with respect to the depth, grain size, and ownership of the personalization.

Keywords

Intelligent tutoring system; personalization; interest; mathematics

1. INTRODUCTION

Adaptive learning technologies that allow for a high level of *personalization* of the learning experience to learners' backgrounds are growing more and more common in classrooms today (e.g., [1]). These technologies allow students in K-12 schools to experience the customization, interaction, and control that they experience when seeking knowledge outside of school [2]. The United States Department of Education [3] recently named the development of personalized approaches to instruction as an important initiative, stating that "The challenge for our education system is to leverage the learning sciences and modern technology to create engaging, relevant, and personalized learning experiences for all learners that mirror students' daily lives and the reality of their futures" (p. X.).

While there is broad consensus that personalized learning is a critical future direction for adaptive learning technologies, there is not yet a detailed knowledge base for effective versus ineffective approaches to personalized learning. For example, personalized learning may need to look different for different subject areas, for different topics being covered within a subject area, and for different types of learners. In addition, there may be important tradeoffs to consider when balancing the time cost of various design decisions for personalized learning with their potential to maximally improve student outcomes.

Here we discuss design tradeoffs relating one particular type of personalization – *context personalization* [4, 5]. We define context personalization as instruction that is matched to students' out-of-school interests in topic areas like sports, movies, or video games. For example, in mathematics, instead of being given a generic problem about a factory producing widgets, a student interested in sports might be presented with a personalized version of the same problem that is instead about scoring goals scored in soccer [5].

This type of personalization involves a number of important design considerations, and in the present paper we report on a study of N = 254 sixth grade students working in the *Reasoning Mind* blended learning mathematics curriculum [6]. Students were randomly assigned to receive different versions of personalized learning materials that varied in the way the personalization was administered by the computer. We examine how different approaches to personalization differentially impacted students' performance and learning

2. LITERATURE REVIEW

2.1 Theoretical Framework

Walkington and Bernacki [7] proposed three primary mechanisms through which context personalization may achieve its effect on student learning. First, personalization may elicit students' *situational interest* [8] in the content to be learned. Interest can in turn promote focus of attention, persistence, and use of learning strategies [8-10]. Second, personalization may promote *utility value* [11] for the content to be learned – by connecting the learning task to students' interest areas, students may begin to see the usefulness and relevance of the learning task to their lives and goals. Finally, personalization may draw upon students' *funds of knowledge* [12]. In mathematics, students may be able to usefully activate their prior knowledge of relationships between quantities to make valid inferences and detect inaccurate reasoning. This has also been referred to as *grounding* [13-14], or the idea that concrete, familiar, and accessible forms of concepts may usefully proceed more abstract forms.

Walkington and Bernacki [7] further discussed how a consideration of these three mechanisms leads to a set of three important design criteria for interventions seeking to personalize learning. The first criteria is a consideration of the *depth* of the personalization intervention – whether the intervention is designed to elicit and draw upon the deep knowledge students have of actually pursuing their interest area, or whether the intervention merely taps into surface features of their interest area (by, for example, simply inserting words into learning tasks related to this interest). The second criteria is a consideration of the *grain size* of the intervention – whether the intervention is intended to be broadly personalized to the interests of all members of a particular age group, school, or class, or whether the intervention truly utilizes adaptive technology to match instruction to an individual's specific and unique interests. The final criteria is a consideration of the *ownership* of the intervention – whether students take an active role in personalizing their own learning, or whether the personalized is imposed upon them by the outside with little student control of the learning experience. We next discuss prior research on personalization using these three criteria as a lens for interpreting research results.

2.2 Past Studies

In one of the most well-known studies of personalization and choice, Cordova and Lepper [4] found that personalizing instruction on order of operations to elementary students' interests (accomplished by replacing words in problem tasks with words students filled in on questionnaires) enhanced learning on a post-test compared to a control condition. They further found that students in conditions where they were given choice over incidental aspects of the problem scenarios performed significantly higher at post-test than those given no choice, suggesting that students' level of ownership is an important factor in promoting learning. Similar findings for the benefit of personalized problems were also reported for arithmetic word problems involving addition and subtraction around the same time period [15-16]. However, more recent studies have shown null effects for personalized learning interventions [17-19].

On the other hand, Walkington [5] conducted a study where Algebra I students received problems in the Cognitive Tutor Algebra intelligent tutoring system. Problems were either written to be personalized to their out-of-school interests (and assigned based on a prior interests survey) or were typical story problems. This study found gains for immediate performance and long-term learning for students who received personalization. Bernacki and Walkington [20] conducted a follow-up study and found that personalized contexts increased algebra students' triggered situational interest and individual interest for algebra. However, they found that only students receiving deeper personalization who initially had low individual interest in mathematics saw long-term learning gains. This echoes other studies that have shown that having students gain an appreciation of the usefulness of learning a content area to their everyday lives, in the form of a utility value intervention, can foster interest and academic achievement for students with low expectations of success and low academic performance [11, 21].

Reber et al. [22] conducted a study of personalized learning in psychology where some students were able to choose psychological scenarios to learn principles from based on their interests, while others received random scenarios from among the interest-based scenarios. This represents a stronger control than that of previous studies reviewed, as it specifically isolates the personalized selection system (rather than the combination of the selection system and the rewritten scenarios that correspond to common interests). With this stronger control, they got the somewhat disappointing result that choosing personalized examples increased interest, but did not impact learning. This study suggests that increased student ownership from problem choice did not impact performance.

Lopez and Sullivan [23] found that both individual- and group-level personalization supported performance over a control group, but that these two conditions were not significantly different from each other. Group-level personalization involved connecting instruction to the interests of groups of students, rather than individual students, thus had a larger grain size.

In more recent work [24], we examined if there were particular broad topics for algebra story problems that adolescents generally performed better on. We found that story problems involving socializing and home and family contexts were generally associated with higher performance than scenarios involve physics, banking, or business, although effect sizes were small. This suggests that personalization at a broad grain size – where students receive problems based on group-level interests rather than receiving problems specifically tailored to individual interests, may still be an effective strategy for improving student learning.

2.3 Current Study

In the present study, sixth grade students received either (1) normal, non-personalized problems in Reasoning Mind, (2) problem variations selected to be matched to their interests based on an interests survey they had responded to earlier, (3) random personalized problem variations not select via a survey, or (4) personalized problems relating to topics of their choice (i.e., they could choose which problem they wanted to solve).

These four conditions varied in the *grain size* of the personalization. In Condition 1, the grain size was very broad – students received problems on whatever topics the authors of Reasoning Mind deemed appropriate. In Condition 3, the grain size was somewhat broad - students received problems written by researchers with an explicit goal of being relevant to the interests of students of this age group, but not problems that were selected to be relevant to their particular interests. In Conditions 2 and 4, the grain size was smaller – students received problems that were personalized to particular topics they expressed interest in either on a survey, or when they selected their own problem. Our first hypothesis (Hypothesis 1) is that smaller grain size conditions (Conditions 2 and 4) will outperform mediocre grain size personalization (Condition 3), who will outperform the larger grain size condition (Condition 1).

The four conditions also varied in the *ownership* of the personalization, or the degree of *choice*. Conditions 1 and 3 had little or no student ownership, as students had no real input into their learning task and its degree of match to their interests. Condition 2 might have a very low level of ownership, if students recalled taking the interests survey and realized that problems were being assigned to them based on their preferences. Condition 4 had a high level of ownership – students took an active role in determining which personalized problem they wanted to solve and thus were able to exhibit control over their learning experiences. Our second hypothesis (Hypothesis 2) is that the condition with a high level of ownership (Condition 4) will outperform conditions with a low level of ownership (Conditions 1, 2, and 3).

Finally, we manipulated the *depth* of the personalization. Students in each of the four conditions received two problem tasks as part of the intervention. We designed one problem task to be shallow and to tap into students' interest area in a surface-level manner only. The other problem task was designed to be deep and describe a specific, detailed, and engaging story about their interest area. Our third hypothesis (Hypothesis 3) is that personalization of the story context will have a larger effect for the deep-level personalized problem task than for the surface-level personalized problem task.

We pursued the following research questions:

1) How do the 4 conditions differ on their immediate performance on the two tasks?

2) How do the 4 conditions differ on measures of long-term learning collected well after solving the two tasks?

We hypothesize that personalization approaches with smaller grain sizes and greater ownership will have the largest effect on performance and learning, because they are optimally designed to elicit interest, enhance utility value, and draw upon prior knowledge. We also hypothesize that greater depth of the personalization will lead to greater performance gains on individual problems.

3. METHOD

3.1 Reasoning Mind Environment

Reasoning Mind is a mathematics blended learning system developed by a nonprofit organization of the same name. Within this system, students study mathematics on computers during their mathematics class time while their teacher is free to conduct targeted interventions with students or groups who are struggling with a concept. The Reasoning Mind organization has developed curricula for second through eighth grades.

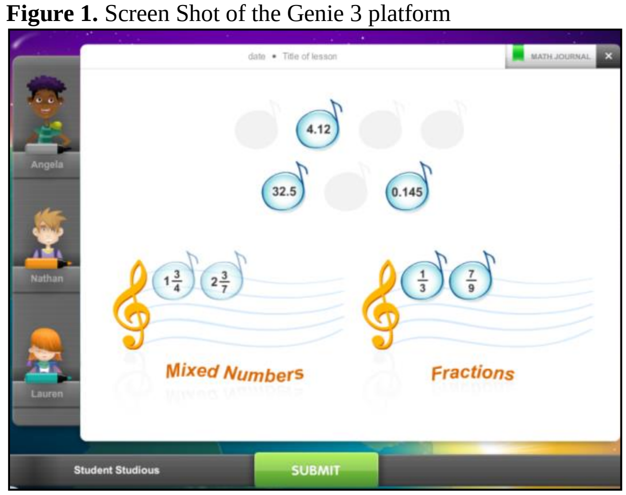
Because the basis of any mathematics curriculum is one of the determining factors in the quality of mathematics education [25], the instructional content of Reasoning Mind grades 2-7 is based on the pedagogical practices of expert Russian teachers using the well-established curricula of Moro and Vilenkin [26-28]. These curricula were developed in the 1970s in the Soviet Union and were an outgrowth of the mathematician Andrey Kolmogorov's reform of mathematics education. These reforms brought to the fore early introduction of algebraic reasoning and techniques, which permeate Moro and Vilenkin [29].

An updated version of the Reasoning Mind system dubbed "Genie 3" was released in 2013 and piloted during the 2013-2014 school year. It was designed to allow for a richer multimedia environment in which the variety of classroom instruction can be simulated faithfully. To that end, the student is immersed in a lesson environment that includes a tutor character, two other student characters, and a virtual blackboard (see Figure 1). The three characters are scripted and speak (narrated by human voices) to each other and to the student, simulating a small tutoring session. They also write on and point to things written on the virtual board with animated markers. The student characters make common mistakes which the real student is asked to correct, they help the real student when he or she gets stuck, and they interact with the real student and the tutor in ways that promote beneficial mathematical attitudes and beliefs.

This rich interactive environment provides a social context which allows students using Genie 3 to engage in the key practices of mathematical work and reasoning embodied in the Common Core Standards for Mathematical Practice [30]. It also embodies research-supported principles of good multimedia design: the use of audio to complement written text and explain visual illustrations [31], instruction coming from socially convincing pedagogical agents using polite, direct, and informal speech [32], use of gestures to help connect auditory and visual information [33].

In the 2013-2014 school year, more than 80,000 students – most in Texas, but also many in California, West Virginia, Tennessee, Louisiana, and other states – used the Reasoning Mind program. Independent evaluations found that the program improves student knowledge of mathematics as measured by state standardized tests, experimenter-created assessments, and the Stanford Achievement Test [34-36]. Quantitative field observations of student affect and engagement have found that students using Reasoning Mind were on task and engaged more often than the average for traditional or other blended learning classes [37-38].

Figure 1. Screen Shot of the Genie 3 platform



3.2 Participants

Participants included 254 sixth-grade students in two small urban middle schools in Texas. There were 132 students in school A split among two teachers and 122 students in school B also split amongst two teachers. 49 percent of the sample identified as male and 51 percent identified as female. The demographic make-up of each school is presented in Table 1.

Table 1. School Demographics and Performance

Demographic Group	% in School A	% in School B
Hispanic/Latino	33.1	97.4
Asian	6.0	0
Black or African American	17.3	1.5
Native Hawaiian or other Pacific Islander	0.4	0
White	36.7	1.1
Two or more races	6.0	0
Economically disadvantaged	69.8	92.2
LEP	10.9	23.5
Special Education	4.8	6.7
Gifted and Talented	5.2	7.1
2014 STAAR Mathematics Passing Rate	30.2	24.3

3.3 Intervention & Measures

Experimental Conditions. Students were assigned to one of 4 conditions: 1) a control condition, 2) a condition where the problem topic is assigned based on students' reported interest on the 4 personalized topics on an interest survey, 3) a condition where students are randomly assigned to one of the four personalized versions of the problem, and 4) a condition where the student is able to choose the problem topic from the 4 personalized topics right before working on the problems. Thirty-one students in the sample did not get assigned to a condition explicitly because they transferred into the intervention classrooms after assignments were made, and thus got control problems. They were not considered for the purposes on this study. Thus our sample size was reduced to 223 students for all analyses, of which 62 were in Condition 1, 52 in Condition 2, 57 in Condition 3, and 52 in Condition 4.

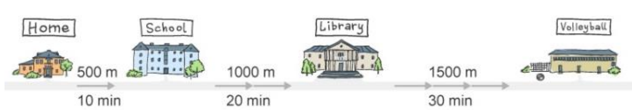
Tasks. Depending on the condition the student was assigned to, they were given one of five versions of the same problem. In each version, the numbers and question remained the same, but the topic of the problem was changed. The control versions of the problems are below:

Q1. *Let's use a table to investigate how the number of books I buy relates to the cost. I was at the bookstore the other day. I bought 1 book, which cost 4 dollars. Then I decided that I wanted two copies of the book. How much will two books cost?*

Number (books)	Cost (dollars)
1	4
4	16

Q2. Kayla walks at a speed of 50 meters per minute. It takes Kayla 10 minutes to walk the distance of 500 meters to school. After school, it takes her 20 minutes to walk the distance of 1000 meters to the library. Her speed did not change. But the time increased. In fact, it increased by a factor of… what? After the library, Kayla walks to volleyball practice. The distance to volleyball practice is times the distance she walks to school. If Kayla's speed did not change, and the distance increased by a factor of 3, then what do you think happened to the time?

Speed	Time	Distance
50 m/min	10 min	500 m
did not change	increased by a factor of 2	increased by a factor of 2
50 m/min	20 min	1000 m
did not change	increased by a factor of 3	increased by a factor of 3
50 m/min	30 min	1500 m



Kayla walks at a speed of 50 m/min

The other four versions of the problems were personalized and changed the topic to sports, shopping, video games, or food. For the first problem (Q1), the personalization was accomplished by simply swapping out the word "books" for another noun – footballs, lollipops, necklaces, and crystals. For the second problem (Q2), the personalization was more involved – each of the locations was replaced with a setting that someone who engaged in the personalized topic would be interested in. For example, the video game variation discussed Kayla traveling to an Enchanted Forest, Dragon Cave, and Wizard's Tower while playing a video game. Readability factors were kept consistent among personalized variations and between personalized and control group problems, as were illustrations. The intervention lesson was given in late April.

Out-of-School Interest Survey. A survey assessing student interest in each of the topic areas was given one month prior to the intervention lesson. Students were asked to rate their level of interest in each topic (0-It's boring, 1-It's okay, 2-I like it, 3-It's my favorite thing), describe how many hours per day they spend on the topic (0-30 minutes, 30 minutes-1 hour, 1-2 hours, 2 hours or more), and rate how much they know about the topic (0-almost nothing, 1-a little, 2-a good bit, 3-a whole lot) on 4-point scales. Problems were chosen for students in Condition 2 based on the topic they said they had the highest level of interest in. If a student rated two topics as equally interesting, the topic they reported spending more time each day on was chosen.

Problem Accuracy. The primary outcome measure is student performance on the intervention problems (Q1 and Q2 above). Student performance was measured with a simple correct or incorrect (0/1) for each part of the problem. Parts of the problem are defined as cells that the student had to fill in while solving the problem scenario that were evaluated by the system for mathematical correctness. The correct/incorrects were then averaged across all problem parts to give a final score on each problem.

Unit Test. A secondary outcome measure was the unit test that students took four days after the intervention lesson. There were 11 lessons and one review lesson covered on the test. There were five required items on the test. Scores on the tests were calculated as a percent correct with each item worth 20 percent (regardless of the number of sub questions in each item). For five students (2 in Condition 1 and 1 in Conditions 2-4), the unit test score was missing and thus is omitted from this analysis. The test questions are below:

1) Find the unknown member of the proportion:
a) $\frac{24}{18} = \frac{8}{x}$; b) $\frac{0.8}{2.4} = \frac{0.9}{x}$; c) $1\frac{1}{7} : x = 2 : 7$;
2) The library got a delivery of 40 books, 35% of which were books by foreign authors. How many books by foreign authors were delivered to the library?
3) At a speed of 90 km/hr, a car travels a distance of 585 km in one day. How many km will this car travel during the same time, if its speed increases to 100 km/hr?
4) 10 monkeys can eat a box of bananas in 18 min. How many monkeys are needed to devour the bananas in 12 minutes?
5) Evaluate the expression: $10*(2.1 \div 3 + 2.7 * 11\frac{1}{9})$

Student Mathematical Knowledge. Although students were randomly assigned to conditions, we used three measures of student mathematical knowledge to control for any enduring systematic differences between groups. First, we controlled for students' accuracy in the "Guided Study" section of Reasoning Mind over the course of the school year. This value ranged from 10-97%. Guided Study is the main study mode in Reasoning Mind. This mode contains the warm-up, presentation of new concepts and ideas, and practice on novel problems. Second, we controlled for students' average level of performance in the lesson prior to the intervention lesson. This lesson covered direct proportions, which consisted of 25 mathematics questions (although not all students were required to solve all problems). Finally, we controlled for students 2014 STARR Test mathematics scores (the state standardized test in Texas), which was also given in April. This value was missing for 10 students (2 in Conditions 1, 2 and 4, and 4 in Condition 3) resulting in a final sample size of 213 students. For 10 additional students, the 2014 STARR test score was not available, but the 2013 score was, and was substituted into the model.

3.4 Analysis Techniques

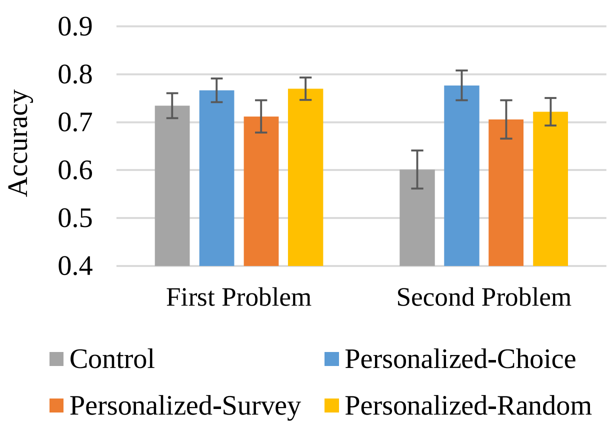
Data were analyzed using mixed effects linear regression models [39] that predicted: percentage of problem parts correct on intervention problem 1 and intervention problem 2 (Research Question 1) and percentage of problems correct on the unit test (Research Question 2). Random effects included Teacher and School. Predictors included Condition (4 levels), STARR Test Score (normalized based on scores of students in the sample), Guided Study Accuracy, and mean percent performance in the previous Reasoning Mind lesson. Gender was tested for inclusion in models, but was never significant. Models were fit using the *lmer()* package [40] in the R software environment. Interactions between Condition and the other predictors were tested for significance using the *anova()* command in R on nested models, which is a χ^2 test that compares model deviance.

4. RESULTS

4.1 Research Question 1: Immediate Performance

Both intervention problems had 5 problem parts with a 6th part that a few students had to complete. The average level of performance across parts of the first problem was 75.0% (SD = 20.0%), and was 70.0% (SD = 28.3%) for the second problem. Figure 2 shows the mean level of performance on the two intervention problems by Condition. Error bars represent standard error of the mean.

Figure 2. Average Performance in Conditions 1-4 on two intervention problems

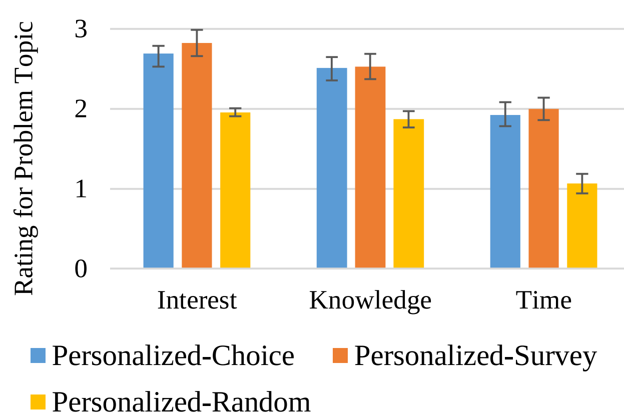


On the first problem (low depth), all conditions performed similarly – there were no performance differences between the control condition and any of the personalization conditions (ps > 0.1) or between any of the personalization conditions (ps > 0.1). Two mathematical knowledge variables, Guided Study Accuracy (B = 0.44, SE = 0.14, p = 0.002) and Percent Performance in Prior Unit (B = 0.33, SE = .08, p < .001) were significant, positive predictors of performance.

On the second problem (high depth), two of the three personalization conditions, Personalization-Choice (B = 15.45, SE = 4.21, p < .001) and Personalization-Survey (B = 9.50, SE = 4.13, p = 0.023) significantly outperformed the Control condition. In addition, Personalization-Random marginally outperformed the Control Condition (B = 7.29, SE = 4.04, p = 0.069), and Personalization-Choice marginally outperformed Personalization-Random (B = 8.06, SE = 4.27, p = 0.061). Once again, both Guided Study Accuracy (B = 0.48, SE = 0.18, p = 0.011) and Percent Performance in Prior Unit (B = 0.49, SE = 0.11, p < .001) were significant positive predictors of performance.

For students in the Personalization-Survey condition (Condition 2), problems had been assigned based on student ratings of their interest in each topic area (sports, video games, food, shopping), and all conditions had also provided ratings for time spent on each topic area and knowledge of each topic area. Figure 3 shows students' interest ratings for the topic areas of the two personalized problems that they actually received in the study. Students in the Personalization-Choice and Personalization-Survey conditions typically received problems for which they had the higher average ratings of interest, time, and knowledge, compared to students in the Personalization-Random. This may in part explain why Personalization-Choice and Personalization-Survey appear to be superior to Personalization-Random in terms of immediate performance on the second problem task. However, neither Interest Rating, nor Knowledge Rating, nor Time Rating were significant predictors of performance on the second problem task (ps > 0.1).

Figure 3. Students in Conditions 2-4 ratings of interest, knowledge, and time (0-3 scale) for problem topics actually received



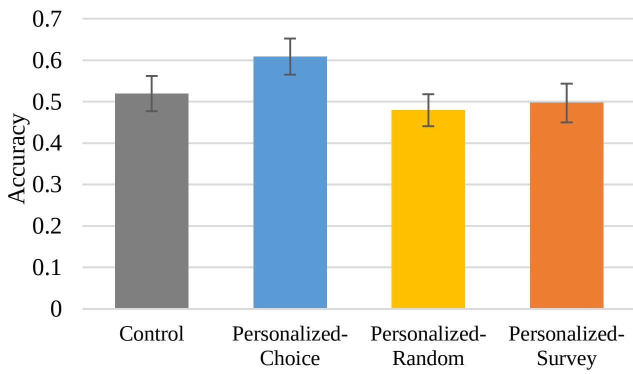
4.2 Research Question 2: Long-Term Learning

The average raw score on the unit test was 49.5% (SD = 32.9%). The only mathematical knowledge variable that was a significant predictor of unit test score was percent accuracy in Guided Study (B = 0.56, SE = 0.27, p = 0.0398).

In terms of Condition (Figure 4), there were no significant differences between the Control Condition (Condition 1) and any of the personalization conditions (Conditions 2-4) in unit test score (ps > 0.1). However, directionally Personalization-Choice scored highest on the unit test, and there were marginally significant differences in performance between Personalization-Choice and Personalization-Random (B = -0.12, SE = 0.06, p = 0.053) and Personalization-Choice and Personalization-Survey (B = -0.10, SE = 0.06, p = 0.087).

This is not surprising given that the intervention involved only one lesson, and the unit test assessed the students' learning from a series of 11 lessons on proportions.

Figure 4. Average Performance on Unit Test in Conditions 1-4



5. DISCUSSION

We conducted a study where N = 213 sixth grade students received either normal story problem or personalized problems selected in different ways in the Reasoning Mind blended learning curriculum. The intervention involved students solving two problems on proportions, and we examined their immediate performance on these problems, their long-term learning as measured by the unit test. We organize our discussion in terms of our three hypotheses regarding grain size (Hypothesis 1), ownership (Hypothesis 2) and depth (Hypothesis 3).

5.1 Hypothesis 1: Grain Size

Hypothesis 1 stated that smaller grain sizes (Conditions 2 and 4) will outperform mediocre grain sizes (Condition 3), who will outperform larger grain sizes (Condition 1).

When looking at measures of immediate performance, we find support for this hypotheses. For students' performance on the second problem task we found that Conditions 2 and 4 (Personalization-Survey and Personalization-Ownership) significantly outperformed Condition 1 (Control). In addition, Condition 3 (Personalization-Random) marginally outperformed Condition 1, and Condition 4 marginally outperformed Condition 3. Thus we see the strongest results when contrasting personalization accomplished through a survey or choice with control versions of story problems. Personalization with a survey and with choice has the smallest grain size, as problems are selected to be relevant to individual learners' interests. On the other hand, the control condition had the largest grain size, as the problems were simply written by curriculum developers to be understandable contexts for sixth grade students.

Looking at measures of long-term learning, we see mixed support for this hypothesis. Condition 4 (Personalization-Ownership) marginally outperformed Condition 3 (Personalization-Random) on the unit test. However, Condition 4 did not outperform Condition 1 (Control), and Condition 2 did not outperform Conditions 1 or 3.

Overall there does seem to be evidence to support Hypothesis 1.

5.2 Hypothesis 2: Ownership

Our second hypothesis stated that conditions with a high level of ownership (Condition 4) will outperform conditions with a low level of ownership (Conditions 1, 2, and 3). As we established differences between Condition 4 and Conditions 1 and 2 in the prior section, the key focus in this section is examining differences between Condition 4 (Personalization-Choice) and Condition 2 (Personalization-Survey).

Examining measures of immediate performance, we did see evidence differentiating Condition 4 from Conditions 1 and 3 on the second task, as described in the previous section. However, we do not see additional evidence that Condition 4 outperformed Condition 2. Examining measures of long-term learning, we see only marginally significant results showing that Condition 4 outperformed Conditions 2 and 3.

Overall there is limited evidence to support Hypothesis 2.

5.3 Hypothesis 3: Depth

Our third hypothesis stated that personalization of the story context will have a larger effect for the deep-level personalized problem than for the surface-level personalized problem. We found strong support for this hypothesis, as none of the personalization conditions (Conditions 2-4) significantly outperformed the control group (Condition 1) on the first low depth problem, but 2 of the 3 personalization conditions (Conditions 2 and 4) did outperform the control group (Condition 1) on the second high depth problem. In addition, the remaining personalization condition (Condition 3) marginally outperformed the control (Condition 1) on the second problem.

Overall there is support for Hypothesis 3.

5.4 Limitations

There were several limitations to the study conducted here. First, a number of students were omitted from the study because they were not assigned to a Condition, and analyses suggest these were the weaker students at the school. In order to get a full picture of how personalization operates across ability levels, it will be important to include these students in future studies.

Second, there may have been order effects in our two personalized problems, as the low depth problem was always presented first and the high depth problem was always presented second. We are currently replicating this study and extending the intervention to encompass four different problems (2 shallow, 2 deep) presented in alternating order. Additional problem tasks will also allow the intervention to perhaps show stronger effects on long-term measures of student learning like the unit test.

Finally, it is critical for personalization studies to examine transformations in students' attitudes towards learning mathematics and their interest in solving math problems as a result of personalization. In our next study, we will include pop-ups that measure students' triggered situational interest in each intervention task, as well as pre-/post- survey measures of their triggered and maintained situational interest in learning mathematics (see [41]) and their perception of the utility value of learning mathematics in terms of how relevant and important math is to their everyday lives (see [11]).

6. CONCLUSION

Adaptive learning technologies like Reasoning Mind that allow for a high level of personalization of the learning experience will continue to become more prevalent in educational settings. It is critical that researchers investigate design principles for these systems, examining outcomes like student performance, learning, and attitudes towards the subject area. Features like personalizing problems to students' particular out-of-school interests can be resource-intensive for developers. Thus determining which aspects of or approaches to context personalization have the most important effect on student outcomes can help designers to focus their design goals and keep the time investment manageable. Studies are also needed that directly compare the effects of a different interest-eliciting interventions within adaptive technology environments, especially when different interventions may involve different levels of designer effort. For example, we are currently conducting studies where we compare the effects of personalization to students' interests to the effects of adding a colorful illustration.

Adaptive environments for K-12 learning have enormous potential to transform education by catering to individual student knowledge and preferences. As we confront a new, digital age where instruction is inexorably linked to online curricular systems, understanding the optimal way to adapt to students' characteristics will become increasingly important.

7. ACKNOWLEDGMENTS

Our thanks to Reasoning Mind, particularly George Khachatryan, for their support of this research.

8. REFERENCES

[1] Fancsali, S. E., Ritter, S., Stamper, J. and Nixon, T. 2013. Toward "Hyper-Personalized" Cognitive Tutors: Non-Cognitive Personalization in the Generalized Intelligent Framework for Tutoring. *AIED 2013 Workshops Proceedings Volume 7*, 71-79.

[2] Collins, A., and Halverson, R. 2009. *Rethinking Education in the Age of Technology: The Digital Revolution and Schooling in America*. New York: Teachers College Press.

[3] U.S. Department of Education, Office of Educational Technology, *Transforming American Education: Learning Powered by Technology*, Washington, D.C., 2010. http://www.ed.gov/sites/default/files/netp2010.pdf

[4] Cordova, D., and Lepper, M. 1996. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*(4), 715-730.

[5] Walkington, C. 2013. Using learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology*, *105*(4), 932-945.

[6] Khachatryan, G. A., Romashov, A. V., Khachatryan, A. R., Gaudino, S. J., Khachatryan, J. M., Guarian, K. R., and Yufa, N. V. 2014. Reasoning Mind Genie 2: An Intelligent Tutoring System as a Vehicle for International Transfer of Instructional Methods in Mathematics. *International Journal of Artificial Intelligence in Education*, *24*(3), 333-382.

[7] Walkington, C., and Bernacki, M. 2014. Motivating students by "personalizing" learning around individual interests: A consideration of theory, design, and implementation issues. In S. Karabenick & T. Urdan (eds.) *Advances in Motivation and Achievement Volume 18* (pp. 139-176), Emerald Group Publishing.

[8] Hidi, S., and Renninger, K. 2006. The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111-127.

[9] Ainley, M., Hidi, S., and Berndorff, D. 2002. Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, *94*(3), 545-561.

[10] Shiefele, U., and Krapp, A. 1996. Topic interest and free recall of expository text. *Learning and Individual Differences*, *8*(2), 141-160.

[11] Hulleman, C. S., and Harackiewicz, J. M. 2009. Promoting interest and performance in high school science classes. *Science*, *326*(5958), 1410–1412.

[12] Civil, M. 2007. Building on community knowledge: An avenue to equity in mathematics education. In N. Nassir. and P. Cobb (Eds.) *Improving access to mathematics: Diversity and equity in the classroom* (pp. 105-117). Teachers College Press.

[13] Goldstone, R., and Son, J. 2005. The transfer of scientific principles using concrete and idealized simulations. *Journal of the Learning Sciences*, *14*(1), 69-110.

[14] Koedinger, K., Alibali, M., and Nathan, M. 2008. Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science*, *32*, 366-397.

[15] Anand, P., and Ross, S. 1987. Using computer-assisted instruction to personalize arithmetic materials for elementary school children. *Journal of Educational Psychology*, *79*(1), 72-78.

[16] Davis-Dorsey, J., Ross, S., and Morrison, G. 1991. The role of rewording and context personalization in the solving of mathematical word problems. *Journal of Educational Psychology*, *83*(1), 61-68.

[17] Bates, E., and Wiest, L. 2004. The impact of personalization of mathematical word problems on student performance. *The Mathematics Educator*, *14*(2), 17-26.

[18] Caker, O., and Simsek, N. 2010. A comparative analysis of computer and paper-based personalization on student achievement. *Computers & Education*, *55*, 1524-1531.

[19] Ku, H., and Sullivan, H. 2000. Personalization of mathematics word problems in Taiwan. *Educational Technology Research and Development*, *48*(3), 49-59.

[20] Bernacki, M., and Walkington, C. 2014. The Impact of a Personalization Intervention for Mathematics on Learning and Non-Cognitive Factors. In Stamper, J., Pardos, Z., Mavrikis, M., McLaren, B.M. (eds.) *Proceedings of the 7th International Conference on Educational Data Mining*, London.

[21] Hulleman, C. S., Godes, O., Hendricks, B. L., and Harackiewicz, J. M. 2010. Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, *102*, 880-895.

[22] Reber, R., Hetland, H., Chen, W., Norman, E., and Kobbeltvedt, T. 2009. Effects of example choice on interest, control, and learning. *Journal of the Learning Sciences*, *18*, 509-548.

[23] Lopez, C. and Sullivan, H. 1992. Effects of personalization of instructional context on the achievement and attitudes of Hispanic students. *Educational Technology Research and Development*, *40*(4), 5-13.

[24] Walkington, C., Clinton, V., Ritter, S., and Nathan, M. (minor revision). How Readability and Topic Incidence Relate to Performance on Mathematics Story Problems in Cognitive Tutor Algebra. *Journal of Educational Psychology*.

[25] Whitehurst, G. J. 2009. Don't Forget Curriculum. Brown Center Letters on Education. Brookings Institution. Retrieved from: http://www.brookings.edu/research/papers/2009/10/14-curriculum-whitehurst

[26] Moro, M.I., Bantova, M.A., and Beltiukova, G.V. 2009. Matematika 2, 3, 4 kl. [Mathematics for the 2nd, 3rd, 4th Grades]. Moscow: Prosveschenie.

[27] Vilenkin, N.Ya. et al. 2007. Matematika. 5 klass: uchebnik dlya obscheobrazovatel'nykh uchrezhdenii [Mathematics. Grade 5: Textbook for General Educational Institutions]. Moscow: Mnemozina.

[28] Vilenkin, N.Ya. et al. 2008. Matematika. 6 klass: uchebnik dlya obscheobrazovatel'nykh uchrezhdenii [Mathematics. Grade 6: Textbook for General Educational Institutions]. Moscow: Mnemozina.

[29] Abramov, A. 2010. Toward a history of mathematics education reform in Soviet schools (1960s–1980s). In B. R. Vogeli (Eds.), *History and World Significance: Russian Mathematics Education* (87-140). Singapore: World Scientific Publishing.

[30] Common Core Standards for Mathematical Practice 2015. retrieved from http://www.corestandards.org/Math/

[31] Fletcher, J.D. and Tobias, S. 2005. The multimedia Principle. In R. E. Mayer (Eds.), *The Cambridge Handbook of Multimedia learning* (117-133). Cambridge, UK: Cambridge University Press.

[32] Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., and Collins, H. 2008. The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies, 66*. 98-112.

[33] Mayer, R.E. 2005. *The Cambridge handbook of multimedia learning*. Cambridge, UK: Cambridge University Press.

[34] Weber, W.A. 2003. An Evaluation of the Reasoning Mind Program at Hogg Middle School. Retrieved from: http://www.reasoningmind.org/pdf/Dr_Weber_Report_2003.pdf

[35] Waxman, H.C. and Houston, W.R. 2008. An Evaluation of the 2006–2007 Reasoning Mind Program. Retrieved from: http://www.reasoningmind.org/pdf/Waxman-Houston_Report_2007.pdf

[36] Waxman, H.C. and Houston, W.R. 2012. Evaluation of the 2010–2011 Reasoning Mind Program in Beaumont ISD. Retrieved from: http://www.reasoningmind.org/pdf/Waxman-Houston_Report_2011.pdf

[37] Ocumpaugh, J., Baker, R. S., Gaudino, S., Labrum, M. J., and Dezendorf, T. 2013. Field Observations of Engagement in Reasoning Mind. *Proceedings of the 16th International Conference on Artificial Intelligence and Education.*

[38] Mulqueeny, K., Mingle, L., Kostyuk, V., Baker, R.S., & Ocumpaugh, J. (under review). Improving engagement in an e-learning environment.

[39] Snijders, T. and Bosker, R. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publications.

[40] Bates, D., and Maechler, M. 2010. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-35. http://CRAN.R-project.org/package=lme4

[41] Linnenbrink-Garcia, L., Durik, A., Conley, A., Barron, K., Tauer, J., Karabenick, S., & Harackiewicz, J. (2010). Measuring situational interest in academic domains. *Educational Psychological Measurement*, *70*, 647-671.