

Using an Intelligent Tutor and Math Fluency Training to Improve Math Performance

Ivon Arroyo¹, James M. Royer,² Beverly Park Woolf¹

¹ Department of Computer Science, ² Department of Psychology
University of Massachusetts, Amherst Mass. 01002
{ivon, bev}@cs.umass.edu; royer@psych.umass.edu

Abstract. This article integrates research in intelligent tutors with studies of basic student competencies, e.g., problem solving strategies and math fluency (the speed needed to retrieve or calculate answers). It describes the impact of computer software designed to improve either strategic behavior or math fluency. Both competencies are key to improved performance and both help decrease the load on working memory as students mature from using physical representations (e.g., counting with fingers) to using cognitive representations. This research evaluated software interventions to improve strategies and fluency and tested their relative magnitude on math post-tutor performance. We discovered that both interventions combined impact math performance and suggest that training in both provide an advantage due to more available memory capacity. Mathematics fluency has an impact on students' cognitive resources that are needed for more difficult (computationally intensive) problems. We suggest that intelligent tutors can be enhanced with math fluidity training activities that help students to make calculations automatically with minimal memory load.

1. Introduction

In this article, we present evidence that suggests that training provided by intelligent tutors for mathematics can be enhanced if students also receive training in *speed* of foundational skill activities --even if those skills are at ceiling accuracy before tutoring starts (i.e. mastered at pretest time). A combination of “fluency training” (the speed with which students either retrieve or calculate answers) and strategical training (approaches to solve specific kinds of problems) should yield higher success at more complex problem solving. We provide theoretical reasons why such a combination of strategy and math fluency training should be potent, based on the fact that problem solving is a two-step process where problems must first be represented in working memory and connected to comparable problem representations stored there, and a second step where calculations are carried out. We suggest that strategy training provided by intelligent tutors currently assist problem representation in the first step of this process: frequent practice assists in the development of stable solution strategies that can be pulled into long-term memory, with hints as a mechanism to help long-term memory searches for stored representations of problems that are similar to the target problem. Tutoring software is still missing though support for the second part of the process, as once a problem has been represented and a possible solution strategy identified, ease at calculations are most likely to affect the correct solution of the problem.

This research is relevant not only because of a novel idea on how to improve tutoring systems’ teaching power in general, but also because of a serious problem in mathematics education in the United States, where students demonstrate a consistent weakness in mathematics performance that becomes magnified over the school years. Additionally great discrepancies in performance exist based on gender and ethnicity. For example, results of the Third International Mathematics and Science Study (TIMSS) involving a half-million students showed that U.S. fourth-graders perform poorly, middle school students worse and high school students are unable to compete. In the 2006 Program for International Student Assessment, US students ranked 17th out of 30 in the science assessment and 24th out of 30 in math. The U.S. National Assessment of Educational Progress (NAEP) categorizes student performance in Mathematics and English into one of four categories: below basic, basic, proficient and advanced.¹ In the recently released 2009 NAEP data only 39% of fourth grade students achieved at the proficient or above level in mathematics, and 18% of those were in the below basic category. The statistics were worse for eighth grade students. Only 34% of those grade 8 students achieved at the proficient or above level and 27% of students at the below proficient level were in the below basic category.

¹ Basic performance denotes "partial mastery of prerequisite knowledge and skills". Below basic performance reflects a skill level below partial mastery of prerequisite knowledge. Proficient performance represents "solid academic performance for each grade assessed." Finally, advanced performance "denotes superior performance in each grade assessed."

The 2009 NAEP tests also indicated discrepancies between the performance of boys and girls and very large discrepancies associated with race and ethnicity. In grade four, 41% of males scored at proficient or above and 18% of male students scored below basic. The females scored slightly worse with 37% scoring at proficient or above and equaling the males at 18% in the below proficient category. In grade 8 performance declined with 36% of males scoring in the proficient or above category and 27% scoring below basic. For females the comparable performance was 32% scoring at the proficient or above level and 29% scoring in the below basic category.

The statistics for racial/ethnic groups showed large differences between groups. In grade four 51% of White students scored in the proficient or above category and only 9% of the students were in the below basic category. This contrasts with 16% of Black students in the proficient or above category and 36% of black students scoring in the below basic category. For Hispanic students, 22% scored in the proficient or above category and 29% scored in the below basic category. The numbers continued to drop for grade 8 students. In grade eight 44% of White students were proficient or above and 17% were below basic. For Black students the comparable statistics were 12% proficient or above and 50% at the below basic level. Hispanic students also were highly discrepant from white students with 17% scoring proficient or above and 43% in the below basic category.

The NAEP data shows an alarming decline in the mathematical performance of students as they move from grade 4 to grade 8, and they show performance differences between males and females and very large differences between racial and ethnic groups. The NAEP data points out the substantial need for more effective mathematics instruction, particularly beyond grade 4 where mathematics performance begins to decline.

The research reported in this article addresses this deficit and is based on research showing that students who display good math strategies do better than students who do not have good strategic performance, and that students who have fluent math fact skills perform better than peers who are slower or less accurate in math fact performance. The goal of the reported research was to determine whether integrating an adaptive tutoring system designed to improve strategic behavior and software for math fact fluency would result in an improvement in mathematics test performance.

1.1. Strategic Ability and Mathematics Performance

As children mature they proceed through a number of stages in their mathematical ability. For example, they shift from physical (e.g., fingers) to cognitive representation of numbers and operations as the load on working memory decreases. The failure to make the shift from physical representation to cognitive representation may signal the presence of a learning disability (Geary, 1994). These stages reflect increasing usage of mathematical strategies. The initial stages involve using concrete representations to count (Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Steffe, von Glasersfeld, Richards, & Cobb, 1983). Further development results in the acquisition

of a mental "number line" to represent numbers and an increase in working memory capacity that allows the cognitive representation of number to replace the need for a physical representation (Poncy, Skinner, & Jaspers, 2007; Case, Okamoto, Griffin, McKeough, Bleiker, Henderson, et al, 1996). The transition from physical representation to cognitive representation of number is a significant accomplishment because it allows students to develop a rich and abstract representation of number, the relationship between numbers, and the development of an understanding of arithmetic operations on numbers (Baroody, Tillikainen, & Tai, 2006; Steffe, Cobb, & von Glaserfeld, 1988).

As children continue to develop mathematical skills they increasingly develop metacognitive abilities, which enable them to know when, why and how to use strategy instruction. These abilities allow them to flexibly use the problem solving strategies they possess (e.g., Montague & Jitendra, 2006). Direct instruction in problem solving strategies has been shown to facilitate the acquisition of strategies and students who have meta-strategic knowledge are more likely to successfully use those strategies (Carr, Alexander, & Folds-Bennett, 1994; Desoete, Roeyers, & Buysse, 2001; Kuhn, Garcia-Mila, Zohar, & Anderson, 1995).

1.2 Computational Fluency and Mathematics Achievement

Computational fluency is another construct that has been shown to be related to mathematics achievement. Fluency refers to the speed with which students can either retrieve or calculate answers to simple mathematics problems. Fast and accurate computation fluency is thought to facilitate mathematical problem solving because it frees up working memory capacity that can then be used in developing problem representations and solution strategies (e.g., Kintsch, 1988).

Computational fluency has been shown to be related to mathematics test performance in grade 2 students (Widaman, Little, & Geary, 1992), and grade 4 students (Kail & Hall, 1997; Whang & Handcock, 1997), and it has been shown to be predictive of mathematical word problem solving performance in seventh and eighth grade students (Zentall, 1990). Royer and his colleagues (Royer, Tronsky, Chan, Jackson, & Marchant, 1999) have also shown that computational fluency is predictive of mathematical test performance in grades 4 through 8 and into college.

The studies mentioned above have largely provided evidence of a correlation between computational fluency and math performance, but intervention research has shown that instruction designed to improve computational fluency also improves overall math performance. Studies showing mathematics improvement as a function of fluency instruction have used computer based interventions (Hasselbring, Goin & Bransford, 1988), audiotaped problems and answers used to prompt correct problem-solution association (Skinner, Bamberg, Smith, & Powell, 1993), Cover, Copy and Compare procedures (Skinner, Turco, Beatty, & Rasavage, 1989), and fluency drills (Royer, et al., 1999) to improve computational fluency and mathematics performance.

1.3 Theoretical Background for This Research

The research reported above indicated that instruction on either strategy use or computational fluency can improve mathematics performance. However, what is not known from prior research is the relative magnitude of the benefit from the two types of interventions. The research reported in this article compares the impact of strategy instruction and fluency instruction separately and in combination. The hypothesis was that the two forms of instruction would complement each another such that a combination of the two would yield greater positive impact than either separately.

A theoretical rationale for the prediction that a combination of strategy and fluency instruction would be particularly potent can be found in the Hummel and Holyoak (1997) theory of analogical access and mapping and in Mayer's (2003) componential theory of mathematical problems solving. The Hummel and Holyoak theory combines connectionist and symbolic features in a general theory of human thinking with specific applications to problem solving, reasoning and learning. The discussion in this article center on the problem solving aspects of the theory.

Hummel and Holyoak propose that problem solvers use previously solved problems stored in long-term memory to assist in solving a current problem. This process involves constructing a representation of the current problem in working memory and then conducting a search of long-term memory for stored problems that can be mapped onto the features of the current problem. If a sufficient match is found the stored representation is also pulled into working memory and the problem solver can proceed with executing steps similar to those used to solve the stored problem in order to solve the current problem.

Strategic behaviors in the kind of problem solving described by Hummel and Holyoak are important for several reasons. First, they can serve to "hone" the problem solving process so that learners don't engage in a lot of superfluous activity while problem solving, thereby creating efficient and elegant problem solving representations to store in long term memory. And second, strategic behaviors encourage the use of efficient search strategies to identify possible solution methods for a current problem.

Mayer (2003) suggested that mathematical problem solving entailed four processes-- translating, integrating, planning and execution. Translation occurs as the problem solver converts the problems into mental representations, while integration involves the merging of these individual representations so as to form a cohesive problem model. Planning and executing deal with the generation and implementation of a solution plan, respectively. As was the case with the Hummel and Holyoak theory, Mayer's theory involves a good deal of interplay between a current problem representation and problem representations stored in long term memory. Mayer particularly emphasizes this point in his distinction between routine and non-routine problems. Routine problems are types that are frequently encountered and that place little burden on the translation and integration processes. Non-routine problems, however, require a considerable amount of memory search to identify possible matches between a problem

representation currently in working memory and one that might be stored in long term memory. This type of search is similar to the search and match process described by Hummel and Holyoak.

The reader should note that problem solving as described by Hummel and Holyoak and Mayer places a heavy burden on working memory capabilities (especially with non-routine problems), and this is where fast and accurate math fact retrieval can produce a positive benefit. As is well known, working memory has severe capacity limitations, both in the number of elements that can be stored at one time, and the period of time that elements can be stored in working memory without constant rehearsal. If a problem solver has to go "off-line" to mentally calculate a math fact or to plug a problem into a calculator there is a chance that the activity may disrupt the problem representation, thereby slowing the problem solving process down, or maybe even disrupting it entirely.

Both of these theories suggest that mathematical problem solving performance should benefit from encouraging students to use strategic behaviors that encourage the development of relevant problem representations, and from practice activities that speed simple mental computations, thereby reducing the load on working memory.

1.4 Combining Strategy Instruction and Computational Fluency Instruction

The strategy interventions used in the present study were implemented in the context of an intelligent tutoring system known as Wayang Outpost. Previous results using the Wayang tutor showed that low performing students who used the tutor improved their learning more than did students of median ability who used the tutor, and they *improved* on more difficult test items (14%) as compared to easy items (5%) while using the tutor (Woolf, Arroyo, Muldner, Burleson, Cooper, Dolan & Christopherson, 2010; Beal, Wallis, Arroyo & Woolf, 2007). The presence of affective pedagogical agents, see Figure 1, was also evaluated. These agents, in the form of a male or female animated cartoon character, provided students with positive feedback and encouragement when problems were answered correctly, and provided encouragement to try again when problems were answered incorrectly. In addition, the agents used the timing of responding to problems to make judgments about the student's affective state. For example, students who were responding fast and inaccurately were given feedback that they should slow down, or perhaps take a short break. Other students who were responding slowly and possibly inaccurately were given a pep talk about trying hard and getting the answers correct.

Earlier research with the Wayang pedagogical agents indicated that women high school students respond to these agents more than do male high school students (Arroyo, Woolf, Burleson, Muldner, Tai, Cooper and Royer, 2010a) and that all students improve their liking of math and their self-concept in math when they use affective pedagogical agents. Female students reported significantly less frustration and more interest (less boredom) when learning companions were used compared to the no learning companion condition and the emotional outcome effects were stronger for females than for males (i.e., while all students improved

confidence and reduced frustration, the results were stronger for females). The version of the Wayang tutor used in this study contained both female and male pedagogical agents. This allowed the opportunity to evaluate whether having a pedagogical agent that was the same sex as the student would make a difference in student learning. This possibility was evaluated by providing half of the students with a same sex agent and half with a different sex agent.

The computational fluency software utilized in the study is commercially available (www.mathsuccesslab.com). Previous research using the Math Success software and procedures like those embedded in the Math Success software has demonstrated that students using the procedures improve math fact fluency, and those improvements transfer to other arithmetic operations and to math problem solving (Royer & Tronsky, 1998)

2. Method

2.1 Participants

Partial data was collected on 237 participants but only 173 of those participants completed all of the tasks in the experiment, and only data from those students with a complete data set will be analyzed. The participants who completed partial data did so for a variety of reasons. Some were absent on one of the days that the experiment was conducted and did not have a complete data set. Others started an experimental session but had to leave the class at some point during the session. Yet others logged onto the computer system incorrectly and contributed data that could not be linked up with other experimental sessions.

The participants with a complete data set were 96 grade 7 students (45 males, 51 females) and 84 grade 8 students (44 males, 40 females).

2.2 Design

The experiment was run separately for the two grades and there were six classes

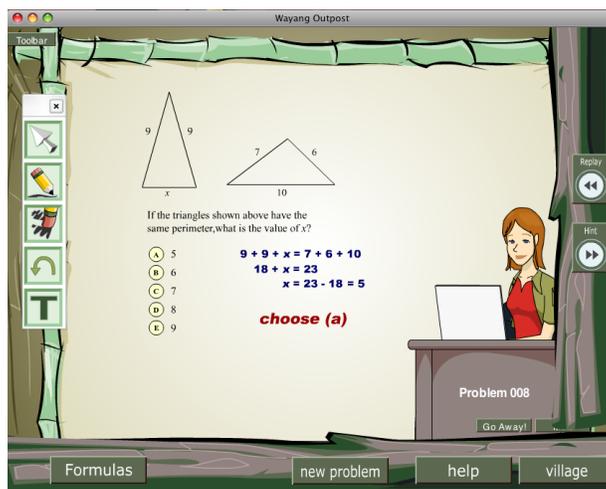


Fig. 1. The Wayang Tutoring System. Gendered affective learning companions talk to students about the need to practice and to exert effort.

participating at each grade. The students in grade 7 had the same math teacher, as did the students in grade 8. Students in the study participated in regularly scheduled math classes while completing the experimental interventions. Several requirements of the experiments such as assisting students in logging onto the Wayang website necessitated that classes be randomly assigned to treatment. Given that there were six classes at each grade level this meant that either two 7th grade classes and an 8th grade class were assigned to one same condition, or two 8th grade classes and one 7th grade class. The distribution of students to treatments is shown in Table 1.

Table 1
Number of Students in Each Treatment Group

<u>Treatment</u>	<u>Grade 7</u>		<u>Grade 8</u>	
	<u>Males</u>	<u>Females</u>	<u>Males</u>	<u>Females</u>
Wayang/Fluency	8	7	14	12
Wayang/ No Fluency	7	8	11	14
No Wayang/Fluency	15	17	8	5
No Wayang/No Fluency	15	17	6	8
Totals	45	49	39	39

2.3 Math Fluency Treatment.

The Math Fluency software provided both *training* and *assessment* of math fact retrieval (MFR) skills. In the training phase, students studied full digital pages of 40 math facts (e.g., two operand addition/subtraction/multiplication/division of at most two digit numbers), and if desired, they could click on each item to hear the answer (to learn or confirm that their answer was right). They then moved to a timing phase by clicking a button on the screen whereupon the page of items would reappear with one of the items highlighted. The students would say the answer to themselves and simultaneously press the space bar which would move the highlight to the next item and the student would repeat the process until all of the items were completed. At the end of the training session the page would reappear and students could again check the correct answer and score any item they thought they answered incorrectly. After scoring the items a graph (shown below) would appear that indicated how long it took the students (seconds per page) to complete the training items, and how accurate they were in responding to the items. This graph would add each practice session to the graph so that the students had a running indicator of their speed and accuracy in completing the math fact items.

In the assessment phase, students were tested for their accuracy and speed of answering a math fact problem. The assessment phase would begin with the appearance of a math fact problem (2 digit addition/subtraction/multiplication/division items) and the student would say the answer aloud and immediately hit the space bar, which would trigger the computer to say the correct answer to the student. Students were instructed to then press the left mouse button if their answer was correct and the right mouse button if the answer was incorrect. Students could cheat in scoring their answers but if they did cheat it would work against the hypothesis being tested in the study. At the end of the assessment session, students could see graphs that showed their speed and accuracy relative to other times they had completed the math fact assessments.

2.4 The Strategy Treatment: the Wayang Intelligent Tutoring System.

Students in the Wayang group used Wayang Outpost, an intelligent multimedia web-based math tutoring software. It teaches students how to solve geometry, statistics and algebra problems of the type that commonly appear on standardized tests by supporting their problem solving (Arroyo, Beal, Murray, Wallis and Woolf, 2004). When using the software, students are presented with mathematics problems, arranged in sequential topics. To answer problems in the Wayang interface, students choose a solution from a list of multiple choice options (see Figure 1). Wayang provides immediate feedback on students' answers by coloring them red or green in the interface. As students work at solving a problem, they ask the computer program for gradual hints that are displayed in a progression from general suggestions to bottom-out solution.

Wayang Outpost makes several pedagogical moves, based on estimates of student abilities and engagement with the program. Wayang provides *cognitive* support via adaptive sequencing of the problems that a student is assigned, depending on students' individual past performance at each problem compared to what is "expected" behavior for each problem (based on estimates of the timing, mistakes and help requested for each problem over thousands of users). Past studies have demonstrated that such adaptive sequencing yields better learning than assigning random problems to students within topics or knowledge units (Arroyo, Mehranian & Woolf, 2010b). In addition, students progress through topics under several different conditions: 1) they have answered all problems in the topic, 2) they have achieved a "mastery" condition – solved enough hard problems correctly, or 3) they have failed enough problems that the content seems above their ability. A spiral organization of the curriculum allows students to revisit topics and attempt problems that they were not initially ready for in previous rounds. This guarantees that students concentrate on problems that are appropriate for them --neither too easy and boring, nor too hard and frustrating (Arroyo & Murray, 2002).

In addition to the cognitive support, the tutoring software provides *meta-cognitive* and *motivational* support, delivered by digital characters designed to act like peers while solving problems. These “learning companions” have two functions. The *motivational* objective is that they speak out messages that would help train students’ attributions for success and failure (e.g., messages encourage students to reflect about myths about innate math abilities, attitudes about problem solving in general, such as reinforcing that effort is a necessary by-product of learning). The learning companions’ motivational messages are tailored to each student’s needs according to a simple assessment of the degree of effort a student invests to develop a problem solution, based on time per action (e.g., whether students are rushing to answer). The *meta-cognitive* purpose of digital characters is to offer general support and advice on how to approach problems (Arroyo, Muldner, Burlison, Woolf & Cooper, 2009) via “strategic” hints or messages (e.g., “How about we try to make a sketch for this problem?”). As an extra meta-cognitive aid, screens with bar charts interleave between problems, helping students reflect about their learning progress. All students using the Wayang software received learning companions, and the gender of the character was assigned randomly. Empirical studies have shown that Progress Charts improve student learning and mastery goal orientation (Arroyo, Ferguson, Johns, Dragon, Meheranian, Fisher, Barto, Mahadevan & Woolf, 2007) and that Learning Companions improve students’ self-confidence and other affective measures within the tutor and after the tutor, especially for lower than median achieving students who incidentally have higher affective needs (Woolf et al., 2010).

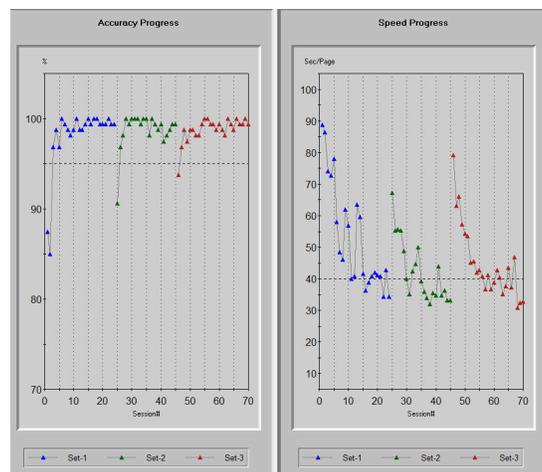


Fig. 2. Student accuracy (left) and speed (right) as displayed in progress charts in the Math Facts Retrieval Training Software

2.5 Dependent variables

All students completed two mock-standardized tests at the beginning and the end of the experiment using online tests presented by the testing module of the Wayang tutoring software. These tests consisted of 10 SAT math problems released by the College Board or released items that were contained in Massachusetts MCAS testing program for grade 8 students. The 10 items

were representative of all the topics taught in the software, and were a mix of easy, medium and hard items. Two tests, A and B, were constructed with similar a similar number of easy, medium and hard items. Some of the questions were “short answer” questions, and some were multiple choice, with both tests having an equal mix of each type of item. The Wayang software collected both accuracy and time to answer information for each test item.

In addition, all students completed a math fact retrieval assessment *at posttest* using the math fluency software. This assessment consisted of 40 items that were an equal mix of two-operand and one-digit by one digit, or two digit by one digit addition, subtraction, multiplication and division items. The software collected both accuracy and speed of answering information for the math fact items.

2.6 Procedure

On the first day of the study all students were directed to the “testing” module of the Wayang software and completed the mathematics mock standardized test (tests were counterbalanced, so that half of students received test A for the pretest and the other half received test B for pretest; the last day tests were reversed for the posttest). Students also completed the pretest math fact retrieval task on the first day of the study. On the last day of the study (day 4) all students completed the math facts retrieval posttest within the Math Fluency software and they completed the mock-standardized test that they had not taken the first day (A or B).

Students in experimental conditions involving the use of the Math Fluency software practiced drilling on single digit multiplication, single digit addition, double and single digit subtraction, and double digit by single digit division, for about 15 minutes every day. Students in experimental conditions involving the use of the Wayang software were directed to the tutoring module (after MFR training in the case of the Wayang-MFR group), where they progressed through 9 topics, practicing each of the problems that were assigned by the adaptive pedagogical module. Before starting to use the Wayang tutoring module, students were encouraged to request hints via the help button and to remember that the goal was to learn from the software.

3. Results

The initial steps in the data analysis were to determine if the two forms of the achievement tests used in the counter-balancing procedure were similar in difficulty and to examine the initial comparability of the four treatment groups. The analysis examining the difficulty of the two test forms was a one way ANOVA where achievement pretest score was the dependent variable and test form was the independent variable. This analysis indicated that test form was not a significant source of variance, $F(1,172) = 1.95$, N.S. Given this result, test form will not be a variable in any of the analyses to follow.

Table 2 shows the means and standard deviations for the pretest achievement test, for the posttest achievement test, for the average amount of time taken to solve each of the problems on the achievement test, and for the average amount of time taken to solve the problems presented

in the math fluency exercises. Note that students in the Wayang/Fluency condition spent less time within the tutoring module of Wayang Outpost (85.8 minutes) than students in the Wayang/NoFluency condition (124 minutes), as students in the former group had part of their class time allocated to the math fluency software.

The initial analyses were designed to determine if the treatment groups were comparable to one another on the initial achievement pretest. The first analysis used achievement pretest score as the dependent variable and grade and treatment as independent variables. The results of this analysis indicated that grade was a significant source of variance, $F(1,165) = 21.9$, $p < .01$, partial Eta squared = .118, but treatment, $F(3,165) = 1.55$, N.S., and the interaction between treatment and grade, $F(3,165) = 1.46$, N.S., were non-significant sources of variance.

The second analysis examining comparability of the treatment groups prior to institution of the treatments took the same form as the above analysis, but used average time to complete the pretest achievement problems as the dependent variable. This analysis indicated that grade, $F(1,165) = .444$, N.S., treatment, $F(3,165) = .277$, N. S., and the interaction between grade and treatment, $F(3,165) = 1.37$, N. S., were all non-significant sources of variance. These results indicate that the treatment groups were performing at a similar level before administration of the treatments.

The next analysis examined one aspect of the question of whether the treatments influenced math performance. Students completing the math fluency training should get faster at solving the math facts presented at posttest. Given the expectation that middle school students would already be at a performance ceiling on relatively simple math facts there was no expectation that exposure to the math fluency training would improve performance accuracy. The means contributing to this analysis are listed in Table 2. The analysis of the time data was an ANOVA that used average time to provide an answer to a math fact as the dependent variable and grade and treatment as independent variables. The results of the analysis indicated that the effect for grade was non-significant, $F(1,165) = .394$, N. S., as was the interaction between grade and group, $F(3,165) = 2.64$, N. S. However, there was a significant effect for treatment, $F(3,165) = 5.16$, $p < .01$, partial Eta squared = .086. Post-hoc contrasts indicated that the Wayang/Fluency group was significantly faster than the Wayang/No Fluency group and the control group, but did not differ from the No Wayang/Fluency group. In addition, the No Wayang/Fluency group was significantly faster than the control group. The analysis of accuracy data on the fluency posttest took the same form as the analysis for time reported above, but used percent correct on the fluency posttest as the dependent variable. The analysis of percent correct performance indicated that grade, treatment, and the interaction between grade and treatment were all non-significant sources of variance. However, the reader will note that an examination of the means presented in Table 2 indicates that groups receiving fluency training did have more accurate performance than groups not receiving fluency training.

Table 2
Overall Means and Standard Deviations () for the Variables in the Study

Variable	Group			
	Wayang/Fluency	Wayang/No Fluency	NoWayang/Fluency	NoWaya/No Flu
Achievement Pretest (%)	65.3 (21.7)	67.9 (21.2)	67.2 (22.0)	63.3 (21.3)
Achievement Posttest (%)	78.0 (21.3)	67.2 (20.5)	61.7 (27.2)	66.3 (18.6)
Adjusted Posttest Achievement Means	80.2 (2.68) ^a	68.4 (2.70)	65.2 (2.71)	70.0 (2.63)
Achievement Pretest Time (seconds)	59.9 (20.1)	61.4 (21.1)	60.4 (16.6)	60.6 (21.8)
Achievement Posttest Time (seconds)	58.7 (22.9)	46.1 (16.4)	58.4 (22.1)	57.3 (23.4)
Adjusted Posttest Time Means	61.3 (3.28) ^a	43.8 (3.31)	61.6 (3.33)	60.9 (3.23)
Fluency Posttest Time (seconds)	1.6 (1.17)	2.4 (1.10)	2.1 (1.28)	2.6 (1.56)
Fluency Posttest Accuracy (%)	96.7 (14.3)	91.9 (18.1)	95.6 (10.1)	94.5 (14.1)
Tutoring minutes in Wayang Outpost	85.8 (18.2)	124 (26.3)	0 ()	0 ()

^a Values in parentheses are standard errors

Having established that the fluency intervention did change math fact retrieval performance, the next set of analyses examined the impact of the treatments on achievement posttest performance. The first analysis was an ANCOVA that used achievement posttest performance as the dependent variable, achievement pretest performance as the covariate, and grade and treatment condition as independent variables. This analysis indicated that grade was a significant source of variance, $F(1,164) = 4.39$, $p < .05$, partial Eta squared = .026, as was treatment condition, $F(3,164) = 5.80$, $p < .01$, partial Eta squared = .096. The interaction between treatment and grade was not significant, $F(3,164) = 2.08$. Post-hoc contrasts between the treatment groups indicated that the Wayang/Fluency group scored significantly higher (at the .05 level) than the remaining three groups on the achievement posttest. None of the other contrasts between treatment groups were significant.

The next analysis examined time taken to solve the achievement test problems as the dependent variable. The means and standard deviations relevant to this analysis are presented in Table 3. Similar to the previous analysis, this analysis used time taken to solve the problems on the posttest as the dependent variable, time taken to solve problems on the pretest as the covariate, and grade and treatment as independent variables. The results of this analysis indicated that grade was not a significant source of variance, $F(1,164) = .965$, N.S., but treatment was significant, $F(3,164) = 6.93$, $p < .01$, partial Eta squared = .112, as was the interaction between treatment and grade, $F(3,165) = 3.12$, $p < .05$, partial Eta squared = .054. Post-hoc contrasts on the treatment groups indicated that the Wayang/No Fluency group solved the problems significantly faster than each of the remaining three groups (at the .05 level). There were no significant differences between the remaining three groups. The nature of the interaction between treatment and grade can be seen in Table 3. The data in Table 3 shows that in both grades 7 and 8 the performance of the Wayang/No Fluency group is considerably faster than the performance of the remaining groups and that the performance of the Control group (No Wayang/No Fluency) and the No Wayang/ Fluency group are comparable in both grades 7 and 8. The group that really changes across the two grades is the Wayang/Fluency group. In grade 7 the Wayang/Fluency treatment seems to slow students down relative to peers whereas in grade 8 the Wayang/Fluency treatment is faster than any group other than the Wayang/No Fluency group. However, it is hard to establish that this effect represents a truly relevant result, given the small cell size when partitioning by grade and treatment (the number of Grade 7 students in the Wayang/Fluency group is $N=7$).

Table 3
Time (seconds) to Solve Math Achievement PostTest Problems
Broken Down by Treatment and Grade

<u>Treatment</u>	<u>Grade 7</u>	<u>Grade 8</u>
<u>Wayang/Fluency</u>		
Mean	67.2 (29.0)	55.6 (21.3)
Adjusted Mean	68.3 (5.2) ^a	54.1 (4.0)
<u>Wayang/ No Fluency</u>		
Mean	42.4 (22.6)	47.0 (13.5)
Adjusted Mean	39.2 (5.3) ^a	48.5 (4.1)
<u>No Wayang/Fluency</u>		
Mean	58.5 (22.5)	64.4 (21.6)
Adjusted Mean	57.6 (3.6) ^a	65.5 (5.6)
<u>No Wayang/No Fluency</u>		
Mean	54.9 (16.0)	65.3 (36.0)
Adjusted Mean	55.9 (3.5) ^a	65.9 (5.4)

The final analysis of central interest in the study was an examination of the impact of the pedagogical agents on achievement test performance. The form of the first analysis was an ANCOVA with achievement posttest as the dependent variable, achievement pretest as the covariate, and grade, treatment and matched or unmatched pedagogical agent as independent variables. As noted in previous analyses, grade and treatment were significant effects in this analysis, and matched or unmatched agent was near significant, $F(1,155) = 3.75$, $p = .054$, partial Eta squared = .024. An examination of the adjusted means for this result indicated that students receiving pedagogical agents that matched their gender performed higher on the achievement posttest (Mean = 73.1) than did students receiving a pedagogical agent that did not match their

gender (Mean = 67.9). A further examination for males and females separately indicated that this advantage was consistent for both males and females. Thus, it appears important to provide middle school students with a character that matches their gender to maximize math learning.

In addition to the analyses of central interest, several other ancillary analyses were performed. The first set of analyses examined the question of whether gender played a role in the outcome of the experiment. One analysis addressed this question using achievement posttest performance as the dependent variable, achievement pretest performance as the covariate, and grade, treatment and gender as independent variables. A second analysis substituted average time taken to solve achievement test problems as the dependent variable and the covariate. Neither of these analyses indicated that gender was a significant main effect, and gender did not contribute to a significant interaction in either analysis.

4. Discussion

The central purpose of the study was to evaluate the hypothesis that strategy training with and intelligent tutoring system combined with math fluency training would produce greater gains in math performance than a control group that did not receive either form of training, and greater gains than groups that received the two forms of training separately. The analysis of achievement pretest data indicated that the treatment groups were equivalent in math performance prior to the implementation of the treatments. The analysis of the posttest data offered support for the central hypothesis in that the group receiving both forms of training scored significantly higher on the posttest than the remaining three groups. In addition, the analyses showed that the fluency training did improve the speed of math fact performance on a posttest of math fact fluency administered at the end of the study. Students in the Wayang/Fluency group answered math facts significantly faster than students in the Wayang/No Fluency group and the control group, but did not differ in math fact speed from students in the No Wayang/Fluency group. The No Wayang/Fluency group also answered math facts significantly faster than the control group.

In addition to achievement test percent correct performance, the analyses also examined time taken to respond to achievement test items. The analysis of this variable indicated that the Wayang/No Fluency group answered achievement posttest questions significantly faster than the remaining three groups, and that the remaining three groups did not differ from one another in speed of problem solution. In addition to the significant main effect for treatment, the speed of problem solution analysis also indicated that there was a significant interaction between grade and treatment. The nature of this interaction, as can be seen in Table 3, was that the relative positions of the Wayang/No Fluency group, the No Wayang/Fluency group and the No Wayang/No Fluency group remained approximately the same across grades. However, the relative position of the Wayang/Fluency group changed markedly across grades. In grade 7 the

Wayang/Fluency group was the slowest of the four groups to reach problem solution. In grade 8, the Wayang/Fluency group was the second to the fastest group to reach problem solution. The possible reasons for this shift will be discussed later.

The final outcome of interest in the study involved the match between the gender of the pedagogical agent and the gender of the student. This analysis indicated that match was a near significant main effect in the analysis ($p = .054$). The nature of the effect was that both males and females scored higher on the achievement posttest if the gender of their pedagogical agent was the same as their own gender. One possible reason for the enhanced performance associated with a gender match between pedagogical agent and student is that the same gender agent provided added motivation for the students to work hard when using the software.

As noted in the introduction to the article, there are good theoretical reasons why a combination of strategy training and math fluency training should be a potent intervention for mathematical problems solving. Both the Mayer (2003) and the Hummel and Holyoak (1997) theories describe problem solving as a multistep process where problems must first be represented in working memory and connected to comparable problem representations stored in long term memory. The problem activated in long term memory then provides a possible solution strategy to the existing problem in that same manner that examining worked problems provide solution strategies to problems currently under consideration. Once the problem has been represented and a possible solution strategy identified, the calculations are performed that result in the solution to the problem.

Strategy training of the type provided by the adaptive tutoring software in this study could assist problem representation in several ways. First, the frequent practice with problems of a particular type assists in the development of stable solution strategies that can then be pulled into long term memory to assist in the solution of newly encountered problems. Second, the hints provided by the software encourage students to perform long term memory searches for stored representations of problems that are similar to the target problem under consideration. Thus the software encourages students to develop the consistent use of memory search strategies when trying to solve a problem that does not immediately activate a solution strategy.

The theories discussed in this article (Hummel & Holyoak, 1997; Kintsch, 1998; Mayer, 2003) also share the attribute of emphasizing the importance of the fact that problem solving takes place in a limited capacity working memory environment. Problem representations are often fragile, particularly if the problems under consideration are not routine problems, and they can fall apart if working memory capacity is stressed. Having to use working memory capacity for the solution of simple math calculations could provide the additional stress that would cause representations to fall apart. Making simple math calculations automatic and virtually capacity free was the intended purpose of the math fluency intervention used in the current study and this

is the likely reason for the beneficial impact of the fluency intervention, but only in combination with an intervention designed to improve problem representation capability.

One interesting result in the present study that is currently unexplained is why the combination of strategy training and fluency training had a differential impact on speed of problem solution in grades 7 and 8. As noted in the results section and at the beginning of this discussion section, the strategy/fluency group was the slowest group to solve posttest problems in grade 7, but was the second fastest group to solve problems in grade 8. At the same time, the fluency/strategy group was the best group in both grades 7 and 8 in terms of percent correct on the posttest. The reason for this outcome is not apparent but probably involves shifting competencies in both problem representation and math fact fluency. Eighth grade students are likely to be more adept and confident about identifying the nature of the problem they are trying to solve after strategy instruction, and they can then quickly move to calculating the answer to a problem, thereby taking advantage of their intervention induced improved math fluency skills. Seventh grade students, in contrast, might be slower and less confident in their ability to identify the nature of the problem they are trying to solve and take longer to complete the problem representation process and delay taking advantage of improved math fluency skills. This explanation is speculative and will require additional research to both replicate and explain the finding.

4. Implications

This research work suggests that tutoring software is more effective at improving students' learning if part of student work is directed to foundational skill activities (i.e., targeting skills that are already mastered) where the focus is speed, beyond accuracy. Students in the experimental group (Wayang/Fluency) spent 30% less time tutoring than the comparison group receiving Wayang Outpost only (Wayang/NoFluency); yet the former group that received fluency and tutoring combined obtained highest achievement at posttest time, and in less time in general. The optimal timing break-down between fluency training and the training of strategy-based skills is still unknown, and a matter of future research. The vision is a new type of tutoring systems with modules for fluency training of already mastered skills, in addition to the traditional tutoring modules where complex problems involving a combination of mastered and unmastered skills are the main target.

It is important not to mis-interpret this study as a fall back to drill and practice kinds of teaching, where rote memorization is of central relevance, or evidence against "conceptual" kinds of teaching, or the teaching of higher order skills and complex problem solving. Instead, this study suggests that part of the training provided by the tutoring software should relate to the speed training of foundational skills—even if already mastered—so as to have them readily

available for higher success at more complex problem solving activities, and higher learning as a result of the experience with the software.

Acknowledgement

This research was funded by a NSF grant, *What kind of Math Software works for Girls?* Arroyo, I. (PI) with Royer, J.M. and Woolf, B.P. (#0734060, HRD GSE/RES), and a grant from the US Department of Education, Institute of Education (IES) *Using Intelligent Tutoring and Universal Design To Customize The Mathematics Curriculum*, to Woolf (PI) with Arroyo, I. and Maloy, B. co-PIs. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

5. References

- Arroyo, I., Woolf, B.P., Burlison, W., Muldner, K., Tai, M., Cooper, D., Royer, J.M., (2010a) Gender Matters: The Impact of Animated Agents on Students' Affect, Behavior and Learning. Technical report UM-CS-2010-020. Computer Science Department, UMASS Amherst
- Arroyo, I., Mehranian, H., Woolf, B. (2010b) Effort-based Tutoring: An Empirical Approach to Intelligent Tutoring. Proceedings of the 3rd International Conference on Educational Data Mining. Pittsburgh, PA.
- Arroyo, I., Woolf, B.P., Royer, J.M., Tai, M.: Affective Gendered Learning Companions, In: 14th International Conference on Artificial Intelligence and Education, pp. 41-48. IOS Press, Amsterdam. (2009)
- Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., Woolf, B.P. (2007) Repairing Disengagement with Non-Invasive Interventions. Proceedings of the 13th International Conference of Artificial Intelligence in Education (AIED07). IOS Press.
- Arroyo, I., Beal, C.; Murray, T.; Walles, R.; and Woolf, B. (2004). Web-based Intelligent Multimedia Tutoring for High Stakes Achievement Tests. In Proceedings of the Seventh International Conference on Intelligent Tutoring Systems 468–477.
- Baroody, A. J., Tiilikainen, S. H., Tai, Y. (2006). The application and development of an addition goal sketch. *Cognition and Instruction*, 24, 123-170.
- Beal, C. R., Walles, R., Arroyo, I., & Woolf, B. P. (2007). Online tutoring for math achievement: A controlled evaluation. *Journal of Interactive Online Learning*, 6, 43-55
- Carr, M., Alexander, J., & Folds-Bennett, T. (1994). Metacognition and mathematics strategy use. *Applied Cognitive Psychology*, 8, 583-595.
- Carr, M., Taasoobshirazi, G., Stroud, R., & Royer, J. M. (in submission). Combined fluency and strategies instruction improves mathematics achievement in early elementary school.
- Case, R., Okamoto, Y., Griffin, S., McKeough, A., Bleiker, C., Henderson, B., & Stephenson, K. M. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development*, 61.
- Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and problem solving in grade 3. *Journal of Learning Disabilities*, 34, 435-447.

- Geary, D. C. (1994). *Children's mathematical development*. Washington, DC: American Psychological Association.
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, 88, 121–151.
- Hasselbring, T. S., Goin, L., & Bransford, J. D. (1988). Developing math automaticity in learning handicapped children: The role of computerized drill and practice. *Focus on Exceptional Children*, 20, 1-7.
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 107,427-466.
- Kail, R., & Hall, L. K. (1997). Direct and indirect influences of developmental changes in processing speed on children's word-problem performance. Paper presented at the Society for Research in Child Development, Washington, DC.
- Kintsch. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.
- Kuhn, D., Garcia-Mila, M., Zohar, A., & Anderson, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, 60(4) #245.
- Mayer, R. E. (2003). Mathematical problem solving. In J. M. Royer (Ed.), *Mathematical Cognition*. (pp. 69-92). Greenwich, CT: Information Age Publishing.
- Murray, T., & Arroyo, I. (2002). Towards measuring and maintaining the zone of proximal development in adaptive instructional systems. *Proceedings of the 6th International Conference on Intelligent Tutoring Systems*. LNCS 2363. Springer-Verlag, pp. 749-758.
- Montague & Jitendra, 2006Montague, M., & Jitendra, A.K. (2006). *Teaching mathematics to middle school students with learning difficulties*. New York: Guilford Press.
- Poncy, B.C., Skinner, C.H., & Jaspers, K.E. (2007). Evaluating and comparing interventions designed to enhance math fact accuracy and fluency: Cover, copy, and compare versus taped problems. *Journal of Behavioral Education*, 16, 27-37.
- Royer, J. M., & Tronsky, L. N. (1998). Addition practice with math disabled students improves subtraction and multiplication performance. In T. E. Scruggs and M. A. Mastropieri (Eds.), *Advances in Learning and Behavioral Disabilities* (Vol 12). Greenwich, Conn.: JAI Press, Inc. (pp 185-218).
- Royer, Tronsky, Chan, Jackson, & Marchant III. (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology*, 24, 181-266.
- Skinner, C. H., Bamberg, H., W., Smith, E. S., & Powell, S. S. (1993). Cognitive, cover, copy, and compare: Subvocal responding to increase rates of accuracy division responding. *Remedial & Special Education*, 14(1), 49-56,
- Skinner, C. H., Turko, T. L., Beatty, K. L., & Rasavage, C. (1989) Cover, copy, and compare: A method for increasing multiplication performance. *School Psychology Review*, 18, 412-420.
- Steffe, L. P., von Glasersfeld, E., Richards, T., & Cobb, P. (1983). *Children's counting types: Philosophy, theory, and application*. New York: Raeger Scientific.

- Whang, P. A., & Hancock, G. R. (1997). Modeling the mathematics achievement of Asian-American elementary students. *Learning and Individual Differences*, 9, 63-88.
- Widaman, K. F., Little, T. D., & Geary, D. C. (1992). Individual differences in that development of skill in mental addition: Internal and external validation of chronometric models. *Learning and Individual Differences*, 4, 167-213.
- Woolf, B.P., Arroyo, I., Muldner, K., Bursleson, W., Cooper, D., Dolan, R., Christopherson, R.M. (2010) The Effect of Motivational Learning Companions on Low-Achieving Students and Students with Learning Disabilities, *Proceedings of the 10th International Conference on Intelligent Tutoring Systems*, 2010, Pittsburgh.
- Zentall, S. S. (1990). Fact-retrieval automatization and math problem solving by learning disabled, attention disordered, and normal adolescents. *Journal of Educational Psychology*, 82, 856-865.