

2005

Effects of web-based tutoring software on math test performance : a look at gender, math-fact retrieval ability, spatial ability and type of help.

Rena L. Walles

University of Massachusetts Amherst

Follow this and additional works at: <https://scholarworks.umass.edu/theses>

Walles, Rena L., "Effects of web-based tutoring software on math test performance : a look at gender, math-fact retrieval ability, spatial ability and type of help." (2005). *Masters Theses 1911 - February 2014*. 2425.

Retrieved from <https://scholarworks.umass.edu/theses/2425>

This thesis is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses 1911 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.



312066 0288 7983 9

EFFECTS OF WEB-BASED TUTORING SOFTWARE ON MATH TEST
PERFORMANCE: A LOOK AT GENDER,
MATH-FACT RETRIEVAL ABILITY, SPATIAL ABILITY AND TYPE OF HELP

A Thesis Presented

by

RENA L. WALLEES

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

February 2005

Psychology


EFFECTS OF WEB-BASED TUTORING SOFTWARE ON MATH TEST
PERFORMANCE: A LOOK AT GENDER,
MATH-FACT RETRIEVAL ABILITY, SPATIAL ABILITY AND TYPE OF HELP

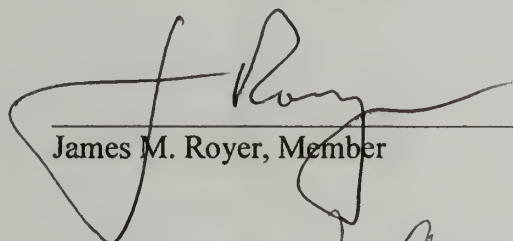
A Thesis Presented

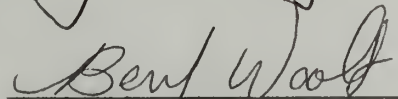
by

RENA L. WALLEs

Approved as to style and content by:


Carole R. Beal, Chair


James M. Royer, Member


Beverly P. Woolf, Member

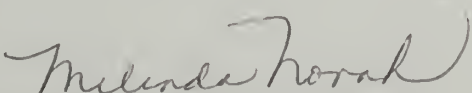

Melinda Novak, Department Head
Psychology

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	v
LIST OF FIGURES.....	vi
CHAPTER	
I. INTRODUCTION.....	1
A. Gender Differences.....	2
B. Potential Predictors.....	5
C. Spatial Cognition.....	8
D. Math-Fact Retrieval.....	11
E. Intelligent Tutoring Systems.....	12
F. Help Strategies.....	15
G. Overview of the Current Study.....	16
II. METHOD.....	19
A. Participants.....	19
B. Materials.....	19
1. Math-Fact Retrieval.....	20
2. Mental Rotation.....	22
3. SAT-Math Pre- and Post-Tests.....	22
4. SAT-Math Tutoring.....	24
5. Information Kiosk.....	25
6. Adventures.....	28
C. Procedure.....	29
D. Scoring.....	31
III. RESULTS.....	33
A. Pre- to Post-Test Performance.....	39
B. Mental Rotation and Math-Fact Retrieval.....	43
C. Predicting Performance.....	46
D. Help Seeking.....	47

IV. DISCUSSION AND CONCLUSIONS.....	52
A. Discussion.....	52
1. Test Performance.....	53
2. Mental Rotation and Math-Fact Retrieval.....	58
3. Predicting Performance.....	60
4. Help Seeking.....	61
B. Conclusions and Future Directions.....	62
REFERENCES.....	65

LIST OF TABLES

Table	Page
1. Experimental Group Pre- and Post-Test Performance for Geometry and Algebra Problems.....	34
2. Control Group Pre- and Post-Test Performance for Geometry and Algebra Problems.....	35
3. Mental Rotation and Math-Fact Retrieval Performance.....	36
4. Help Seeking Performance by Help Mode.....	38
5. Experimental Group Pre- and Post-Test Performance by Gender and Mode.....	44
6. Help Seeking Performance by Gender and Mode.....	48

LIST OF FIGURES

Figure	Page
1. Math-fact retrieval task.....	21
2. Mental rotation task.....	23
3. Example of analytic help.....	26
4. Example of visual help.....	27
5. Pre- to post-test improvement by question type and group.....	40
6. Pre- and post-test scores by answer category.....	42
7. Pre- to post-test improvement by pre-test rating.....	45
8. Pre- to post-test improvement by help rating.....	50

CHAPTER I

INTRODUCTION

High-stakes achievement exams are becoming the gatekeepers of academic and career opportunity for today's students. In many states students must pass achievement exams, such as the MCAS exam in Massachusetts, in addition to required coursework in order to receive a high school diploma. The majority of students hoping to gain admission to college must take the SAT. These exams are appearing at all major junctions in a student's career. In order to succeed in school, and perhaps even in life, students must excel on these pivotal exams.

Differential performance by males and females on high-stakes mathematics achievement exams has received much attention, both in the academic literature and in the popular forum. The discrepancy in performance, with males systematically outperforming females, is a serious issue in education that has far-reaching and long-lasting effects. For example, students who receive lower scores on the SAT-Math exam have a reduced chance of gaining entrance into prestigious schools and desired math-intensive majors, such as computer science and engineering (Webb, Lubinski & Benbow, 2002). SAT scores profoundly influence the probability of receiving scholarships, such as the National Merit Scholarship. Lower scores may also reduce self-esteem and turn students away from math-oriented fields of study (Eccles, 1997). Performance on high-stakes exams clearly impacts future opportunities for both male and female students.

In an attempt to improve student performance, several computer-based mathematics tutoring systems, such as Carnegie Learning's software, have been

created and disseminated. Wayang Outpost is one such web-based intelligent tutoring system that aims to prepare students for the SAT-Math exam. The system is unique in that it was created with the female user in mind. Mathematics adventures featuring female characters, which require students to solve multi-step real-world problems, were created with the intention of drawing in female students and making the material more engaging for everyone. The Wayang tutor encourages the adoption of new ways of thinking about mathematics problem solving, and provides detailed help in order to allow students to succeed on even the most difficult problems.

When attempting to tutor students, it is essential to understand individual characteristics that will affect the usefulness of the help given. An in-depth analysis of how exactly the gender difference manifests itself and what other factors may come into play is necessary. The literature points to two factors, in conjunction with gender, that could potentially impact test performance: math-fact retrieval ability and mental rotation ability. One goal of this thesis is to gauge the veracity of these variables as predictors of math test performance and to determine how best to tutor students of various backgrounds and ability levels.

A. Gender Differences

Gender differences in math test performance are a popular topic of inquiry. The sheer quantity of research in the area is indicative of its vast importance for students, teachers and scientists. As described earlier, performance on these exams has several practical implications for test takers. Students who do not score well on the SAT are less likely to gain entrance into top schools. Poor performance by females relative to males may lower females' self-esteem and push them away from

math-oriented programs of study. Similarly, teachers are under the intense scrutiny of school administrators, parents, and the government. They may feel pressure to “teach to the test” and perhaps overlook other important course material. Finally, researchers are often interested in the more theoretical issue of understanding exactly how and why the gender disparity occurs and finding innovative ways to close the performance gap.

Past research has uncovered several important characteristics of the gender disparity that continue to inform current research. First, the gender difference was found to be most pronounced in clusters of the highest performing students, defined here as those students scoring between 700 and 800 on the math portion of the SAT (Stanley & Benbow, 1983). These talented students are likely to be applying to competitive colleges and programs and therefore taking exams such as the SAT, making effective tutoring all the more important for this group.

Second, from a young age, males are found to outperform their female classmates on standardized exams (Benbow, 1988). This weakens the claim that males outperform females simply because they have taken more math courses. It appears as though something more basic, perhaps even something genetic, underlies the gender disparity.

Finally, females have been found to outperform males in mathematics courses from early elementary school all the way into graduate work (Kimball, 1989). This is a well-known finding in experimental psychology, as well as a general pattern often recognized by educators themselves. Several reasons for this disparity, many of them

social in nature, have been provided. However, further analysis is necessary before making any solid conclusions.

More current research, such as that of Bielinski and Davison (1998), focuses on the test itself in an attempt to discover problem characteristics that may lead to differential performance by males and females. Interestingly, the researchers found that males perform better on multiple-choice items, algebra problems, real-life problems, and problems of high cognitive complexity. Females, on the other hand, perform better on free response items and textbook-like items, which are typically of lesser difficulty. Bielinski and Davison suggest that the performance difference may be caused by a subtle shift in the abilities measured as test items increase in difficulty. Less difficult questions may tap more low-level reasoning while more difficult questions may rely more heavily on spatial visualization, a skill at which males often outperform their female counterparts (Linn & Peterson, 1985).

This finding may also help to clarify the achievement exam to classroom performance disparity. If females excel at textbook-like problems while males prefer real-world problems, this pattern of performance would be expected instead of surprising.

Hyde, Fennema, and Lamon (1990) conducted a meta-analysis in order to organize the plethora of information available on gender differences in math test performance. The study included data from over 100 studies and over three million participants. The general finding was that the difference, with males outperforming females, is most apparent in highly selective samples of high school students. Another critical finding was that the difference appears to be diminishing, with studies

published in 1973 and earlier showing significantly larger differences than studies published in 1974 and later.

Hyde and colleagues (1990) warn that the gender difference, though relatively small, appears in critical areas and therefore should not be ignored. They stress the importance of preparing female high school students for careers in math-related fields through an intensive problem solving curriculum. They implore teachers and administrators to provide the necessary materials for women to succeed in mathematics.

Despite the progress that has been made in the area, more research is necessary to further explain the gender disparity. In addition to explanation, steps must be taken in order to close the gender gap. In order for that to happen, researchers must be able to pinpoint the factors that underlie the gender difference.

B. Potential Predictors

Research in the field supports a male advantage on high-stakes mathematics achievement exams. The next step is to determine what factors account for this variation. Researchers have explored several possible predictors, most of which fall into three general categories: social factors, biological factors, and systematic test bias.

Support has been found for both social and biological explanations, and it is likely that there is no clear-cut answer but rather a mix of environmental and genetic influence. The social factors include math anxiety, math self-confidence, and internal values, while the biological hypothesis focuses on genetics and hormones. The test bias hypothesis has received some popular support but has not enjoyed systematic support in the academic literature.

Beal (1999) highlights areas of discontent with all three hypotheses. The test bias hypothesis is simply not supported by the extensive research on high stakes math achievement exams (Willingham & Cole, 1997). Researchers have gone to great lengths to explore each test question for potential bias and rewrite questions accordingly, but still the gender gap remains. Item bias does not seem to be a strong enough factor to account for the gender discrepancy (Beal, 1999).

Instead of test bias, it may be the nature of the testing environment that is particularly disruptive to female performance (Beal, 1999). Females may be more negatively affected than males by the emphasis on time constraints and the importance placed on a single test. However, this explanation breaks down when one considers the verbal portion of the SAT exam, where females do not perform more poorly than their male counterparts. If females were in fact negatively affected by the nature of the testing environment, the detrimental effects would be seen in the overall score not just in the mathematics portion.

Problems also exist for biological hypotheses. For example, it is unclear why spatial ability would increase performance on an exam such as the SAT but not on classroom activities (Beal, 1999). Perhaps the answer lies in the skills necessary to solve problems on each type of exam. Further research is necessary in order to clear up inconsistencies and explore viable hypotheses.

Clearly each hypothesis has its downfalls and no single hypothesis can account for all of the variation. Therefore, research such as that conducted by Casey, Nutta^{ll}, and Pezaris (1997), which focuses on both social and biological predictors, is useful in determining independent and joint contributions. The researchers looked at spatial

ability and internalized beliefs, and found that SAT-Math performance was mediated by both mental rotation ability and math self-confidence. Gender differences for math anxiety, however, did not significantly affect performance. Interestingly, spatial ability was found to be almost twice as influential as math self-confidence. Spatial ability is an often studied ability that has the potential to be an important predictor of performance.

Despite having received relatively little experimental attention, math-fact retrieval is another potential predictor of interest. The term refers to the ability to quickly and accurately retrieve math facts from memory. Research by Royer, Tronsky, Chan, Jackson, and Marchant (1999) has shown gender differences in math-fact retrieval and has indicated that it may be an alternative to spatial cognition or perhaps an ability that works in conjunction with spatial ability. Math-fact retrieval could help to explain the discrepancy between the male advantage on test performance and the female advantage on in-class performance. More research is necessary, however, before solidifying this claim. Math-fact retrieval is an attractive potential predictor of performance, and perhaps part of the gender disparity puzzle.

The current research focused on two of the aforementioned abilities: spatial cognition and math-fact retrieval. The importance of other environmental and social factors should not be underestimated, but were not measured directly in this study. One goal was to clarify the possible relationship between spatial ability and math-fact retrieval ability, and to determine how they are related to gender and test performance, among other variables.

C. Spatial Cognition

Previous research has shown that males from select populations outperform females on high stakes math achievement tests such as the SAT (Stanley & Benbow, 1983). In-depth analyses of test questions reveal that males have an advantage on the most difficult questions, which may require spatial visualization for successful completion (Bielinski & Davison, 1998). Accordingly, one proposed reason for the gender discrepancy is that males have superior spatial ability, and that this difference may increase their performance relative to females on difficult items. This alleged relationship has been examined by several researchers.

Numerous meta-analyses have been conducted in order to examine gender differences in spatial ability (Hyde, 1981; Linn & Peterson, 1985; Voyer, Voyer & Bryden, 1995). Establishing gender differences in various measures of spatial reasoning is an important first step. If experimental research is incapable of finding gender differences in basic spatial abilities, it would be unreasonable to continue to view spatial ability as an underlying cause of the gender discrepancy in math test performance.

Linn and Peterson's (1985) meta-analytic study found gender differences for certain measures of spatial ability but not for others. The greatest gender differences were found for mental rotation, with males significantly outperforming females. Similarly, Voyer and colleagues (1995) found differences only on certain tests of spatial ability, such as mental rotation. Both of these meta-analyses included data from hundreds of studies, and attest to the existence of gender differences in mental rotation, the spatial ability of interest in the current study.

Subsequent researchers probed specific relationships between measures of spatial ability and test performance. For example, Reuhkala (2001) explored the relationship between visuo-spatial working memory and mathematical ability. She found that 39% of the variance on the math ability test was explained by the mental rotation task. In a second experiment, Reuhkala confirmed her previous results while simultaneously showing that other working memory components, the central executive and the phonological loop, did not significantly correlate with math ability. Reuhkala reasoned that deficits to visuo-spatial working memory may diminish the “space” allocated for solving difficult problems, thus lowering scores on math achievement exams.

Manger and Eikeland (1998) took a slightly different approach and first examined test problems in order to find out the skills necessary to solve them. They found that the most difficult questions necessitated the use of spatial tactics as opposed to purely analytic knowledge. Additionally, the researchers found a male advantage on the most difficult test problems, but no gender discrepancy on easier questions. It appears as though the male advantage may stem from male superiority on the most difficult questions which require spatial proficiency. However, the authors note that there were still significant gender differences after spatial ability was accounted for, leaving room for additional predictors.

While many studies confirm gender differences in spatial ability, specifically mental rotation, few give any suggestions as to whether or not this difference can be alleviated. In a meta-analysis performed by Baenninger and Newcombe (1989), it was found that spatial ability performance could be improved through training. However,

this increase was found for both males and females, so the training did not succeed in closing the gender gap. This is an encouraging finding given that the current study attempted to tutor students in certain spatial skills including estimation and visualization. It is possible that extensive tutoring within the framework of an intelligent tutoring system will succeed in lessening the gap.

Similarly, Subrahmanyam and Greenfield (1994) have suggested that increased video game usage may contribute to superior spatial ability. Their main finding was that practicing spatial skills in a video game environment improved spatial ability irrespective of gender for those who were initially poor at spatial tasks. If spatial ability is important in mathematics performance and can be improved through experience, the current study's goal of improving spatial skills through tutoring help seems attainable.

Some researchers suggest that spatial ability is such an integral part of overall intelligence that testing students for this ability would complement and enhance traditional measures of intelligence (Shea, Lubinski & Benbow, 2001). Research points to this ability as a gatekeeper to careers in many scientific disciplines, attesting to its overall importance to students.

Shea and colleagues (2001) performed a longitudinal study to evaluate the utility of measuring spatial cognition for predicting outcomes such as educational success and occupational choice. They found that verbal, quantitative, and spatial ability were all correlated with educational and occupational outcomes. In fact, they found spatial ability to be an even better predictor than quantitative ability. Spatial

ability is clearly an important construct that deserves increased attention in future research.

D. Math-Fact Retrieval

The spatial cognition hypothesis has received support in previous research, but it does not appear to be the only factor involved. Casey, Nuttall, Pezaris, and Benbow (1995) found that for females only 12% of the variance in math aptitude was explained by mental rotation ability. Furthermore, some researchers claim it does not affect math achievement at all (Weiner & Robinson, 1986).

Math-fact retrieval is a second potential predictor that may account for a portion of the unexplained variance. Royer, Tronsky, Chan, and colleagues (1999) conducted nine experiments in order to explore various effects of math-fact retrieval on math test performance. Three studies demonstrated that math-fact retrieval predicted performance on achievement tests for students of various ages. Three additional studies showed that males and females perform differently on tests of math-fact retrieval, whereby the fastest males outperform the fastest females in measures of both speed and accuracy. The final three studies examined gender differences for retrieval tasks in general, finding that females slightly outperformed males on verbal processing speed tasks but no differences on general retrieval tasks.

The math-fact retrieval hypothesis is appealing because it is able to account for phenomena other theories cannot explain. For example, it may be able to account for the finding that females outperform males in the classroom, but not on timed achievement tests. In the classroom, females are more likely to have adequate time to finish problems, while on achievement exams there is often intense time pressure. If

males are indeed faster at retrieving math facts, they are gaining precious fractions of a second with each problem solved. This adds up over time and may leave males with additional time to answer problems. Also, students who are faster at math-fact retrieval to the point where it is automatic may be able to free up cognitive resources necessary in answering complex problems (Royer, Tronsky, Chan, et al., 1999).

Critics of the math-fact retrieval hypothesis believe the theory cannot account for gender differences in problems that do not require arithmetic, such as geometry problems (Geary, 1999). Other critics claim that Royer, Tronsky, Chan, and colleagues did not directly test their hypothesis, that the magnitude of the gender difference is too small to be of importance, and that math-fact retrieval surely cannot be the major mechanism driving the gender difference (Wigfield & Byrnes, 1999).

Royer, Tronsky, Marchant, and Jackson (1999) accept some of the criticisms, while attempting to refute others. They concede that more research is necessary in order to verify the reported findings. However, they maintain that math-fact retrieval does have immense potential in adding to the corpus of knowledge on gender differences in high-stakes math test performance. Further research will likely clarify these issues.

E. Intelligent Tutoring Systems

Research on intelligent tutoring systems has suggested that there are important features inherent in the programs themselves that may either help or hinder students' learning. For example, the style of interaction that a student has with a computerized tutor is incredibly important in either reducing or increasing the anxiety that students, and women in particular, feel while in a learning situation (Cooper & Stone, 1996).

Furthermore, the pedagogical agent's voice conveys important social cues, and these cues can affect learning. Social agency theory says that the use of a human voice in a multimedia program activates the social conversation schema, thus making students more likely to engage in deep cognitive processing (Mayer, Sobko & Mauton, 2003).

Wayang, the intelligent tutoring system that is the focus of this study, was created with the female user in mind, aiming to create a non-threatening, engaging and educational environment. Several steps were taken in the creation of the tutor in order to assure its attractiveness to female users. For example, the three mathematics adventures have leading female researchers in the field of orangutan study as the main characters. The voice used to convey the help in the tutor is that of a woman. Great concern was taken in assuring the environment would be hospitable to females, while also engaging to male students.

Critics argue that web-based tutors are not nearly as effective as one-on-one human tutoring, but research has shown that deep learning can occur through the use of intelligent tutoring systems. Moreno, Mayer, Spires, and Lester (2001) found that students who interacted with a pedagogical agent performed significantly better on tests of transfer and were more interested in the program than students who did not see a pedagogical agent. As classroom size grows and teachers become less able to reach each individual student, intelligent tutoring systems will continue to grow as important tools in motivating students and sparking an interest in learning.

Even with the introduction of effective technologies into the classroom, it is unclear whether these programs will be fully implemented and utilized. Cuban, Kirkpatrick, and Peck (2001) offered and confirmed through their research two

reasons why this may occur. First, teachers simply do not have the time to locate and assess software. With growing duties for teachers as well as swelling classroom size, teachers are unable to allocate time to searching for educational technologies. Second, training for the usage of available technologies is not always offered at convenient times.

Easy to use, web-based programs like the Wayang Outpost alleviate many of these issues and may thereby increase classroom usage of intelligent tutoring systems. Teachers would need little to no computer expertise, as Wayang is run completely from one website. Time would not be wasted searching for software and subsequently downloading it onto a whole classroom full of computers. Wayang is a creative, viable solution to the aforementioned problems.

Wayang has other advantages as well. Previous studies on the influence of math-fact retrieval and mental rotation on performance have largely relied on outcome measures of mathematics skills. A major advantage to intelligent tutoring systems is that detailed assessments of performance are readily available. For example, data is collected on a problem-by-problem basis, so that latency and incorrect choices can be analyzed in addition to a simple overall score. Also, Wayang and other tutoring systems are able to determine usefulness of type of help while keeping all other variables constant. Students in both the analytic and the visual help groups are immersed in the same environment, simply viewing different types of hints. The advantages to using intelligent tutoring systems are numerous, including ease of use for teachers and effective and engaging learning tools for students.

F. Help Strategies

Research has shown that gender differences do not appear for all types of math problems, and that some topics are associated with more substantial differences than others (Rosser, 1989). Specifically, Rosser reported that the largest gender differences on SAT-Math exams have been found for geometry problems. The Wayang tutor was created to be comprehensive, but also to maintain a focus on geometry problems due to their prominent role in the gender divide.

Casey and colleagues (1997) note that many math problems can be solved in a variety of ways. For example, students may use a textbook-like approach whereby they assign names to variables and create equations. This is termed the analytic approach in the current study. Other students may employ a more visual approach, using angle estimation and visualization in order to quickly and accurately solve a problem. Some students may be able to effortlessly switch between the two approaches given the nature of the question on which they are working.

Students are often taught the analytic approach in the classroom, but not many teachers explicitly instruct students on the visual approach to problem solving. This may be where males are gaining an advantage over females. If males are naturally superior to females in spatial ability, which is to be determined, they will have more flexibility when it comes to solving these complex problems.

This difference would be most apparent when students are working on high-stakes exams. Classroom tests are created by the teachers and are therefore assessing the student's ability to solve problems in the analytic manner in which the material was taught. High-stakes exams, on the other hand, are often assessing higher level

skills and the student's ability to transfer knowledge to more challenging problems. Having a wider variety of problem solving strategies available would be quite beneficial in this type of environment.

Geometry problems are especially amenable to this breakdown in help strategies, analytic versus visual, which is another reason Wayang focuses on this type of question. Casey and colleagues (1997) point out that it is important to see which types of intervention will maximally assist females, which is one of the goals of the current study. Wayang provides students with new ways of looking at problems with the goal of increasing their strategy flexibility. This increased flexibility is expected to contribute to an improvement in test scores.

Additionally, Gallagher (1992) reported that most items on the SAT-Math exam that favored males required insight, while most items that favored females required algorithmic solutions. Again, this coincides with previous research and with the finding that females outperform males in the classroom but not on achievement exams. The Wayang tutor provides both types of help, and this study aimed to clarify which type of students, based on gender, math-fact retrieval ability, and mental rotation ability, would benefit most from each type of help.

G. Overview of the Current Study

The current study is an extension of an experiment run in the spring of 2003 at the University of Massachusetts. Participants were high school students from two schools in Western Massachusetts who used the Wayang system for several class periods. Results from this study were quite encouraging. Students were shown to benefit significantly from the tutoring program, as indicated by a 22% increase in

scores from pre- to post-test. Students also significantly reduced the number of skipped questions, further increasing their overall scores.

While no differences were found with respect to help mode, visual or analytic, several gender differences were seen. Though all students improved from pre- to post-test, males outperformed their female counterparts on both exams. The tutor was not effective in reducing the gender gap. With respect to help seeking, females were more likely to ask for help, while males were more likely to make several incorrect attempts.

The current study used an updated version of the Wayang Outpost, which was expanded to include a revised test of mental rotation ability, a new online math-fact retrieval task, an updated battery of test problems, and web-based pre- and post-tests, among other more technical features. Two new adventures were added as well in order to further test transfer of skills learned in the tutor. The general look and feel of Wayang was updated in order to be more engaging and age-appropriate for high school students.

The main goal of the current study was to test the veracity of Wayang Outpost with high school students preparing for the SAT-Math exam. Potential gender differences in pre- and post-test performance, math-fact retrieval, and mental rotation were gauged, and Wayang's effectiveness in increasing student performance was determined.

A second goal, new to the current study, was to determine the most effective way to tutor students of various abilities. Composite scores were computed for math-fact retrieval and mental rotation respectively in order to gauge individual student

ability. Numerous tests were then performed in order to discover the benefits and drawbacks of the two distinct types of help. Additionally, the predictive ability of math-fact retrieval and mental rotation were assessed.

CHAPTER II

METHOD

A. Participants

Two hundred and eighteen students from two culturally diverse high schools in Western Massachusetts participated in this study. There were 94 males and 124 females, ranging in grade level from sophomores to seniors. One class from each of the two schools was randomly assigned to the control condition. In all, 149 students were assigned to the experimental condition and 69 students were assigned to the control condition. The study was conducted in the Spring of 2004.

B. Materials

The Wayang Outpost tutoring system was created at the University of Massachusetts by an inter-disciplinary team of Psychologists and Computer Scientists. The system is available entirely on the web, allowing easy access to any student whose school or home has internet capabilities. The web-based format also allows for frequent updates to the system, an especially important capability in the constantly changing area of mathematics education.

The overall atmosphere Wayang attempts to create is an integral element of the system. Students are intended to feel as though they are in an animated classroom in a remote hut on the island of Java. They are exposed to Indonesian culture, renowned researchers, and exciting math adventures where their problem solving ability helps save baby orangutans and curb illegal logging, among other feats. Most importantly, help is available at every juncture in order to facilitate the learning process.

Wayang Outpost combines personalized SAT tutoring with real-life adventures in order to hone mathematics skills that will translate into higher scores on high-stakes mathematics achievement exams. By encouraging new ways of thinking and fostering the use of effective existing strategies, Wayang has the ability to capitalize on student strengths while facilitating areas of weakness.

The Wayang system is divided into several components, most of which were utilized in the current study. Online math-fact retrieval and mental rotation tasks, two SAT exams, an SAT tutor, three mathematics adventures, and an information kiosk are all available to student users. Based on a student's progression through the system, however, certain choices may be temporarily unavailable. For example, students must have completed the math-fact retrieval task before they move on to the tutor or the adventures. The server updates the user's information so that only appropriate choices are available at any given time.

In the current study, students completed the math-fact retrieval task, the mental rotation task, SAT-Math tutoring, and three adventures. Students also completed paper-and-pencil versions of the pre- and post-test and visited the information kiosk. Each task is described below.

1. Math-Fact Retrieval

The math-fact retrieval task is a verification version of the original task utilized by Royer, Tronsky, Chan, and colleagues (1999) which requires students to respond via microphone. In the on-line version used in this study, the student viewed each problem presented on the screen and then chose whether the given statement (e.g., $7+2=10$) was true or false (see Figure 1). Students indicated their response by

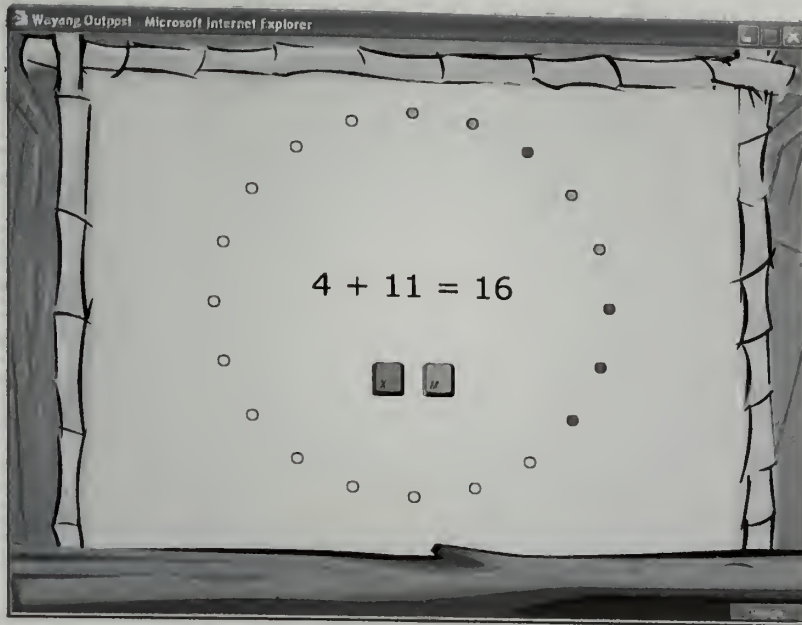


Figure 1. Math-fact retrieval task.

pressing a pre-determined key representing their answer. A correct response was indicated on-screen by a green circle, while an incorrect response was indicated by a red circle. Students completed 20 randomized trials, which represented a mix of addition, subtraction, multiplication, and division. Due to prior research indicating a lack of motivation, students were allowed to view their accuracy via the dot system as well as being continually aware of how many problems remained. At the end of the task, summary statistics appeared showing overall accuracy and mean time.

2. Mental Rotation

The mental rotation task is a verification task, similar in structure to the math-fact retrieval task. Students completed 20 randomized trials; on each try they were shown two images and had to decide if they matched (see Figure 2). The images are digital versions of the original Vandenberg (1978) print images showing geometric figures in various orientations. Students judged whether the two figures were the same, i.e., if one could be rotated to match the other. Keystroke responses indicating same or different were recorded by the server. Students were provided with feedback regarding accuracy, via the red and green circles, and how many more problems were left to complete, indicated by the clear circles. Summary statistics appeared upon completion of the task.

3. SAT-Math Pre- and Post-Tests

Despite the availability of online pre- and post-tests, paper-and-pencil tests were used in the current study. In previous evaluations of the Wayang system where online tests were utilized, students were examined in crowded computer labs which

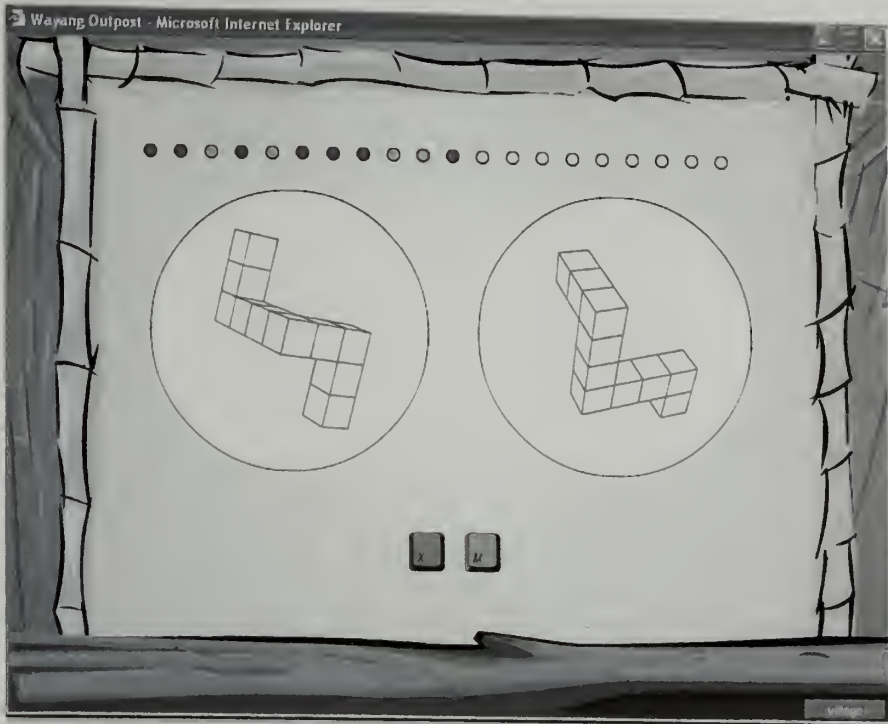


Figure 2. Mental rotation task.

left little room for working out problems on scrap paper and fostered excessive and unwanted student interaction. Until updates are made to the system to allow for easier access to virtual scrap paper and online calculators, it was determined that the best results would come from paper-and-pencil exams.

Each student completed a paper-and-pencil pre-test before using the tutoring module and a paper-and-pencil post-test upon completion of the tutor. The pre- and post-tests were constructed from problems taken from previously administered SAT-Math exams (Green & Wolf, 1998; Princeton Review, 2000). There were two forms, Form A and Form B, each consisting of 21 problems. Problems were selected according to their difficulty level, as provided by the College Board, and the skills necessary to successfully complete them, as determined by the experimental team. Six questions on each test were algebra problems, and the remaining 15 questions were geometry problems. The Wayang tutor provided specific tutoring for the geometry skills but no tutoring for the algebra skills, allowing for a within subjects control. Half of the students received Form A as the pre-test and half received Form B as the pre-test. Students received the alternate form as the post-test. The two forms were found to be of comparable difficulty in a previous study with high school students.

4. SAT-Math Tutoring

After students completed the SAT pre-test, they entered the SAT tutoring module. Students were presented with an SAT-Math problem, including a graphic and five answer choices, modeled after the original SAT problem. Students could choose to attempt to answer, or to receive help by clicking a button. Hints appeared one by

one as the student requested them. Students were able to choose an answer at any time.

There are two versions of the SAT tutor – one that provides mainly visual help and one that provides mostly analytic help. The visual help showed the student a solution to the problem that involved strategies such as estimation and visualization. In contrast, the analytic help showed the student an algorithmic or computational solution. An example of an algorithmic solution is shown in Figure 3 and an example of a visual solution is shown in Figure 4. Students were randomly assigned to receive either visual or analytic help.

There are 60 problems available in the tutoring module, and students had two sessions of approximately 40 minutes each to work on the problems. The number and length of sessions varied as a function of the individual school and classroom schedule. The exact number of problems that each student completed in the tutoring module varied because some students worked more quickly than others, and those who requested help may have completed fewer problems than those who did not request help. The total time students spent working with the tutoring module was automatically recorded by the system.

5. Information Kiosk

The information kiosk had several functions within the Wayang system. First, it was a place where students could learn about orangutans. An orangutan character, Theodore, spoke to the student while showing a virtual slideshow containing various pieces of information on the species. A second section highlighted three leading female researchers, Anne Russon, Lori Perkins, and Elizabeth Fox, who starred in the

Wayang Outpost Microsoft Internet Explorer

problem: Problem 2 hint: solve_for_x

In rectangle $ABCD$, diagonal AC makes a 30° with side ID . If $AC = 10$, what is the area of the rectangle?

$25\sqrt{2}$
 $25\sqrt{3}$
 48
 50
 100

$A = h \times w$
 What do you know about the sides of a 30-60-90 triangle?
 $A = x \times \sqrt{3}x$

Find the value of x
 $2x = 10$
 Solve for x .
 $x = \frac{10}{2}$
 $x = 5$

Figure 3. Example of analytic help.

Wayang Outpost Microsoft Internet Explorer

problem: Problem 6 hint: solve_estimation

In the figure above, what is the value of a ?

- A 10
- B 20
- C 28
- D 36
- E 45

$3a$ is about 90°
 $3a \approx 90$
 $a \approx 90 / 3$

error messages help volume

Figure 4. Example of visual help.

adventures. Students were able to learn about the women within the adventures they were about to see and to gain a better understanding of the system as a whole.

6. Adventures

There are currently three adventures in the Wayang system. They are an integral part of the overall structure, and for many students they are the most interesting and engaging component. Each adventure consists of several multi-step problems, usually four or five, that must be completed in order to achieve some overarching goal. The difficulty level of these problems is generally higher than that of the problems seen in the tutor. In addition, when a student asks for help within the adventure, they are directed to a problem within the tutor that requires similar problem solving skills. In other words, the help is not nearly as direct or comprehensive as in the tutor. Therefore, student interaction is encouraged while working within the adventures. The purpose of these tasks is to provide students with real-world problems that can be successfully completed with knowledge gained from the SAT tutor.

The first adventure is known as the “save the baby orangutan” adventure. Dr. Anne Russon, a renowned orangutan researcher, is in need of help. There has been a fire and a baby orangutan is deserted in a remote section of the island. The student must help Anne solve several problems, such as determining the shortest route from one point to another, in order to find the baby and take him to a nearby hospital.

The second adventure is the illegal logging adventure. Lori Perkins, a proponent of orangutan safety and well-being, has found trucks illegally taking logs

and thus depleting the environment. The student must help Lori solve problems in order to curb the illegal logging and save the orangutans' habitat.

In the final adventure, students must help Dr. Elizabeth Fox rebuild an orangutan nursery that has been badly damaged in a storm. Students must determine the length of the wood to use in building the foundation, calculate how many tiles are needed for the roof, and complete other necessary tasks using their math skills.

C. Procedure

The procedure for the experimental group was as follows. The first session occurred in a typical classroom setting, with no researchers present. Students were administered a paper-and-pencil pre-test proctored by their mathematics teacher. They were given 30 minutes to complete the test, and no further assessments were completed that day.

The second session occurred in an internet equipped computer lab at the students' respective school. Students were given handouts explaining the day's activities including necessary web sites. Researchers guided students on how to create usernames and passwords. Once everyone logged in, students completed the online math-fact retrieval task and the online mental rotation task. Students had a maximum of five minutes to complete each task, and no student failed to complete either task within the time limits. Students were then instructed to proceed to the SAT-Math tutoring hut. They worked within the tutoring system for the remainder of the class period.

The third day of testing, occurring in the computer lab, commenced with students logging back in to the Wayang system. Students were instructed to re-enter

the SAT-Math tutoring module and to work on additional problems. They were allowed to work within the tutoring module until the end of the class period or until they completed all 60 problems.

Students worked with the SAT tutor for approximately 40 minutes each day. Some saw visual hints and some saw analytic hints. Controlling for gender, students were randomly assigned to either the visual or analytic help condition. Specifically, approximately half of the females received visual help and half received analytic help, and the same occurred for males.

During the fourth day of testing, students were instructed to enter the information kiosk, which highlights, among other things, the characters used in the adventures and an informational orangutan slideshow. After having viewed the background information, students entered the SAT-Math adventures. Students were allowed to complete as many adventures as possible within the class period. There was no specified order to the adventures, so students may have viewed different ones than their classmates. Most students, however, did complete all three within the given time period. At the end of the day, or upon completion of the adventures, students were given a link to the Wayang website so they could return at their leisure and further explore the system.

The fifth day of testing was completed in a typical classroom setting with only the mathematics teacher present. Students were administered the paper-and-pencil post-test exam proctored by their teacher. They were given 30 minutes to complete the test, thus completing the Wayang test battery.

Students in the control group were administered the paper-and-pencil pre-test in the same manner as students in the experimental group. They did not, however, use the system prior to the post-test. The paper-and-pencil post-test was administered four days after the pre-test in a manner identical to that of the experimental group. In the interim students received the normal mathematics curriculum decided upon by their respective teacher. Once both the pre- and post-test were complete, students were allowed to explore the Wayang system.

D. Scoring

Pre- and post-tests were hand scored by the experimenter. Raw scores as well as percentages were made available for analysis in terms of correct answers, incorrect answers, and skipped questions. A scoring system similar to the one used in the SAT was utilized to account for guessing: three points were given for each correct answer, one point was taken away for each incorrect answer, and .2 points was subtracted for each unanswered question (College Board, 2004). A composite score was computed for the pre-test and a median split was used in order to assign students to one of two groups, low scorer or high scorer, to be used in further analyses.

The mental rotation and math-fact retrieval task data was collected and stored electronically. The server kept a record of each individual student response, as well as the time taken to make the response. Composite scores were computed for each task, rewarding above average accuracy while taking away points for above average speed.

A plethora of information was available concerning interaction within the tutor and help seeking behavior. Access to information, including which problems in the tutor were seen, how many attempts were made to answer the problem, number of

hints seen per problem, minutes spent in the system, and the answer that was ultimately chosen, were all available for analysis. New variables were computed in order to determine when help was seen with respect to when an answer was given. Additionally, a median split was used to divide students into two groups based on their help seeking behavior.

Unfortunately, the data from the adventures was not accurately collected by the server. Therefore, no analyses were computed for the adventures. However, students did complete surveys that offered valuable information that will inform future versions of the system.

CHAPTER III

RESULTS

A major goal of the current study was to examine student performance within Wayang Output, a web-based SAT-Math tutoring system. In order to achieve this goal, in-depth analyses were performed with respect to pre- and post-test performance, behavior within the math-fact retrieval and mental rotation modules, and help-seeking within the tutoring module. Later analyses include major independent variables such as gender and help mode.

Pre- and post-test performance of the experimental group (see Table 1) and the control group (see Table 2) was captured in a variety of ways. Problems were broken down into two types: geometry problems which were specifically tutored in the Wayang system and algebra problems which were not tutored. There were 15 geometry problems and six algebra problems. Students were scored on number of problems correctly answered, number of incorrect problems, and skipped questions. Percentages were computed for number of correct answers, skipped questions, and incorrect responses by dividing the respective values by the total number of questions. A composite score was computed whereby students earned three points for every correct answer, lost one point for every incorrect answer, and lost one-fifth of a point for each skipped question. This scoring style was based upon the scoring system of the SAT (College Board, 2004).

The mental rotation task and the math-fact retrieval task were scored in a similar manner. Table 3 summarizes data for both of these tasks. The server collected

Table 1

Experimental Group Pre- and Post-Test Performance for Geometry and Algebra Problems

Performance	Pre-Test		Post-Test	
	Geometry	Algebra	Geometry	Algebra
# Correct	3.47 (2.09)	1.38 (1.25)	4.46 (2.32)	1.59 (1.19)
# Incorrect	9.71 (3.50)	3.89 (1.65)	9.92 (2.76)	4.08 (1.47)
# Skipped	1.81 (3.50)	.74 (1.57)	.62 (1.66)	.33 (.99)
% Correct	23.16 (13.91)	22.96 (20.85)	29.73 (15.49)	26.54 (19.77)
% Incorrect	64.74 (23.32)	64.81 (27.44)	66.12 (18.47)	68.02 (24.43)
% Skipped	12.10 (23.30)	12.35 (26.23)	4.15 (11.07)	5.43 (16.52)
Composite	.35 (7.98)	-.10 (4.72)	3.34 (9.30)	.63 (4.73)

Note. Standard deviations are in parentheses.

Table 2

Control Group Pre- and Post-Test Performance for Geometry and Algebra Problems

Performance	Pre-Test		Post-Test	
	Geometry	Algebra	Geometry	Algebra
# Correct	5.18 (2.32)	1.43 (1.43)	4.97 (2.92)	1.66 (1.26)
# Incorrect	8.25 (3.41)	3.35 (1.86)	8.80 (3.38)	3.92 (1.55)
# Skipped	1.57 (2.60)	1.35 (1.96)	1.22 (2.13)	.42 (1.09)
% Correct	34.50 (15.46)	23.75 (23.84)	33.17 (19.48)	27.72 (20.95)
% Incorrect	55.00 (22.73)	55.83 (31.02)	58.67 (20.50)	65.36 (23.88)
% Skipped	10.50 (17.34)	22.50 (32.59)	8.17 (14.20)	6.93 (18.21)
Composite	6.96 (9.42)	.66 (5.52)	5.88 (11.58)	2.10 (5.57)

Note. Standard deviations are in parentheses.

Table 3

Mental Rotation and Math-Fact Retrieval Performance

Performance	Mental Rotation	Math-Fact Retrieval
# Problems	33.99 (35.00)	30.41 (22.11)
# Correct	21.20 (19.34)	26.85 (19.43)
Time/Prob.	5217.36 (4243.07)	4158.10 (2552.66)
% Correct	65.35 (15.38)	88.90 (10.46)
Composite	-.01 (.75)	-.04 (.63)

Note. Standard deviations are in parentheses.

data on the number of problems viewed, the number of correctly answered items, and the average time taken, measured in milliseconds, to complete one problem. Percent correct was computed by dividing the number of correct items by the total number of items. Composite scores were calculated by first determining the averages for both percent correct and time per problem. An equation was then created whereby the individual student's average time per problem was divided by the average time per problem for the entire sample, and the resulting value was subtracted from the student's average percent correct divided by the average percent correct for the entire sample. This equation, inspired in part by Royer, Tronsky, Chan, and colleagues' (1999) dual focus on speed and accuracy, was designed to reward above average accuracy while detracting points for above average time.

Several measures of student behavior were collected concerning interaction with the help provided in the tutoring module (see Table 4). The total number of hints viewed and the number of mode-specific hints were collected by the server. Mode-specific hints refer to hints that could be classified as distinctly "visual" or "analytic". The first few hints of most problems were generic, so if students only asked for one hint on each problem they may have never seen hints specific to their tutor mode.

The total number of problems where help was seen, the average number of hints viewed within these problems, the average time spent in help, and the average number of incorrect attempts in problems were also collected from the server data. Several variables were then computed in order to determine when an answer was given with respect to when help was seen. Specifically, the number of problems where help was seen before the correct response, the number of problems where help

Table 4

Help Seeking Performance by Help Mode

Performance	Analytic	Visual
Total Hints	69.09 (98.49)	79.12 (101.92)
Specific Hints	7.11 (6.84)	11.10 (11.12)
# Problems	12.25 (8.32)	15.22 (10.26)
Hints/Prob.	2.90 (1.09)	2.64 (1.01)
Time/Help	86.60 (58.44)	86.52 (66.66)
Incorrect Attempts	1.15 (.64)	1.09 (.67)
Help Before Answer	11.25 (8.47)	14.23 (10.38)
Help Before Attempt	10.32 (8.09)	12.72 (10.22)
All Help/Answer	10.75 (8.38)	13.90 (10.32)
All Help/Attempt	9.45 (7.54)	12.03 (10.06)

Note. Standard deviations are in parentheses.

was seen before any attempt, the number of problems where all hints were seen before the correct answer, and the number of problems where all hints were seen before any attempt were computed for later analysis.

Together, these variables give a general overview of student interaction within the Wayang system. Further analyses reveal possible effects of both within-subjects and between-subjects variables, such as gender, mode of help, and math-fact retrieval ability.

A. Pre- to Post-Test Performance

One goal of this study was to determine if students would improve from pre- to post-test after working within the Wayang tutoring system. An ANOVA was performed with condition (experimental/control) as a between subjects factor and test (pre/post) and type of question (geometry/algebra) as repeated measures. There was a main effect of condition, $F(1, 173) = 9.36, p = .00$, with the control group ($M = 3.90$) outperforming the experimental group ($M = 1.10$). There was also a main effect of question type, $F(1, 173) = 32.39, p = .00$, with higher scores on geometry problems ($M = 4.13$) than algebra problems ($M = .87$). There was no significant main effect of test. However, there was a significant two-way interaction between type of question and experimental condition, $F(1, 173) = 9.68, p = .00$, as well as a significant three-way interaction between test, type of question, and experimental condition, $F(1, 173) = 5.97, p = .02$ (see Figure 5).

Paired sample comparisons revealed that the only significant change from pre- to post-test was for the experimental group with geometry problems, $t(134) = -3.74, p = .00, (M = .35$ and $M = 3.34$, respectively). The experimental group did not

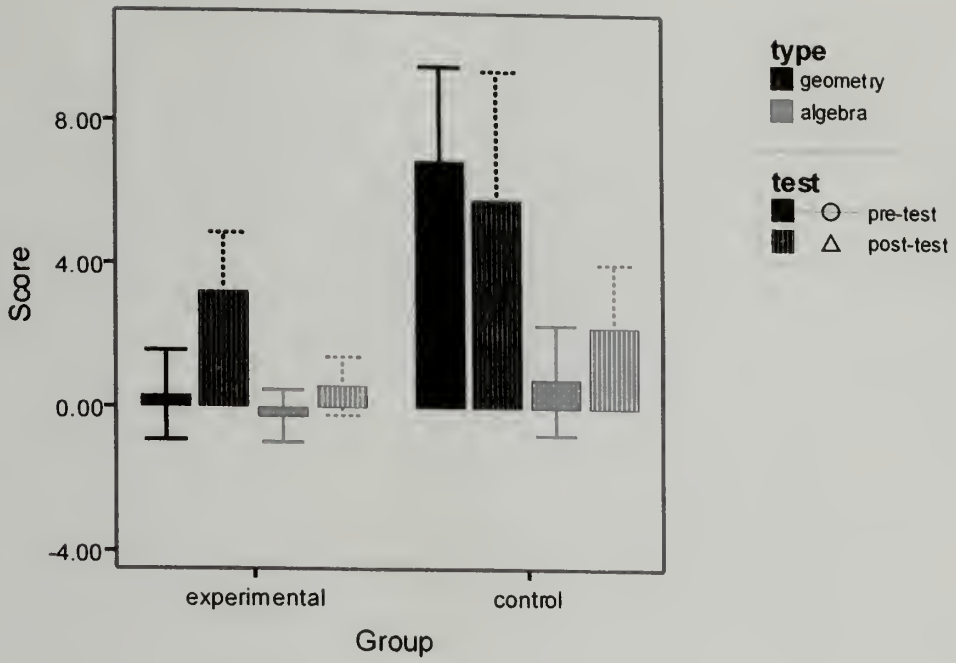


Figure 5. Pre- to post-test improvement by question type and group.

significantly improve on the non-tutored algebra problems, and the control group did not improve from pre- to post-test on either question type. For this reason all future analyses were performed on the experimental group for tutored problems only.

Students who used the system were expected to increase in the percentage of correct questions while decreasing in the percentage of skipped and incorrect questions from pre- to post-test. A repeated measures ANOVA was performed with test (pre/post) and response category (correct, incorrect, skipped) as repeated measures. There was a significant interaction between test and response category, $F(2, 133) = 17.26, p = .00$ (see Figure 6). Paired sample comparisons revealed a significant increase in the percentage of correct questions from pre- to post-test, $t(134) = -4.91, p = .00, (M = 23.16$ and $M = 29.73$, respectively), a significant decrease in the percentage of skipped questions, $t(134) = 3.98, p = .00, (M = 12.10$ and $M = 4.15)$, and no significant change in the percentage of incorrect questions ($M = 64.74$ and $M = 66.12$).

Sub-goals included determining whether gender and mode of help, either visual or analytic, would interact with test performance. An ANOVA was performed for the experimental group with gender and type of help (visual/analytic) as between subjects factors and test score (pre/post) as a repeated measure. As indicated previously, there was a main effect of test performance, $F(1, 107) = 12.91, p = .00$, demonstrating an increase in score from pre- ($M = .35$) to post-test ($M = 3.34$). No significant main effect for type of help was found, nor were there any significant interactions. However, there was a marginally significant main effect of gender, $F(1, 107) = 3.71, p = .06$, with males ($M = 3.75$) outperforming females ($M = .91$).

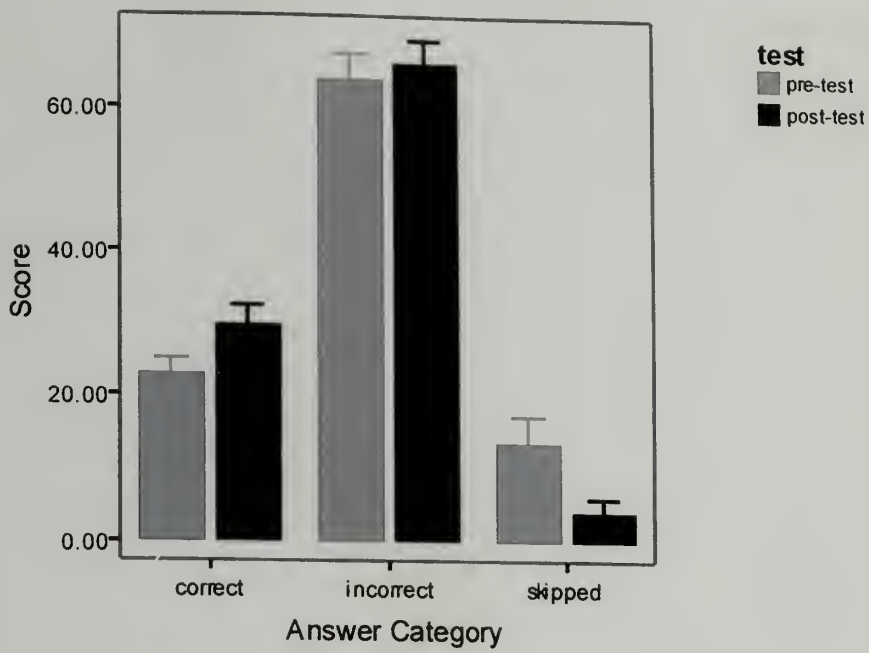


Figure 6. Pre- and post-test scores by answer category.

The magnitude of this difference may have been obscured by large within group variability. Experimental group performance is summarized in Table 5.

In order to gauge Wayang's effectiveness in tutoring students of varying ability, students were divided into groups based on their pre-test scores using a median split technique. An ANOVA was conducted with pre-test rating (low scorer/high scorer) as a between subjects measure and test performance (pre/post) as a repeated measure. As would be expected, there was a significant main effect of pre-test rating, $F(1, 133) = 58.59, p = .00$, with those who scored well on the pre-test ($M = 5.96$) significantly outperforming those who did poorly on the pre-test ($M = -2.10$).

However, there was also a significant interaction between pre-test rating and pre- to post-test performance, $F(1, 133) = 29.33, p = .00$. Paired sample comparisons revealed that students who performed poorly on the pre-test showed significant improvement from pre- ($M = -5.52$) to post- test ($M = 1.31$), $t(68) = -6.76, p = .00$, while those who received higher scores on the pre-test showed no significant change from pre- ($M = 6.48$) to post-test ($M = 5.49$). The system appears to be most effective for those who start off at a disadvantage (see Figure 7).

B. Mental Rotation and Math-Fact Retrieval

Previous research suggested that a gender difference favoring males would be found in mental rotation accuracy as well as math-fact retrieval speed. A multivariate ANOVA was performed with gender as the independent variable and mental rotation accuracy, time per problem, and composite score as the dependent variables. No significant differences were found. A separate multivariate ANOVA was performed with gender as the independent variable and math-fact retrieval accuracy, time per

Table 5

Experimental Group Pre- and Post-Test Performance by Gender and Mode

Performance	Male		Female	
	Visual	Analytic	Visual	Analytic
Score				
Pre-	2.36 (8.17)	2.65 (11.30)	-2.13 (6.64)	-.26 (7.37)
Post-	4.82 (11.07)	5.15 (9.15)	3.63 (9.92)	2.38 (7.44)

Note. Standard deviations are in parentheses.

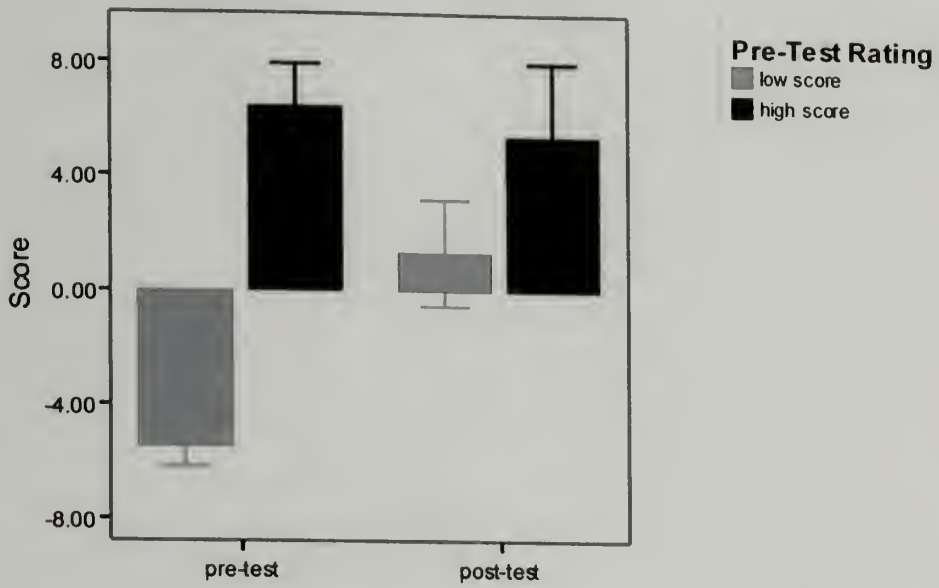


Figure 7. Pre- to post-test improvement by pre-test rating.

problem, and composite score as the dependent variables. Females ($M = 90.71$) significantly outperformed their male counterparts ($M = 87.02$) on the measure of accuracy, $F(1, 155) = 5.15, p = .03$. A marginally significant gender difference was found for composite score, $F(1, 155) = 2.58, p = .10$, with females ($M = .06$) outperforming their male peers ($M = -.10$).

In order to explore the predictive ability of the math-fact retrieval and mental rotation tasks, simple correlations were run between pre- to post-test change score and several measures: percent correct, time per problem, and composite score. No significant correlations were found. Regression analyses indicated that including percent correct and time per problem in addition to the composite score did not add predictive power, with R values remaining the same in each case, .11 for math-fact retrieval and .50 for mental rotation. Neither math-fact retrieval score nor mental rotation score significantly predicted change in score from pre- to post-test while not holding any other variables constant.

C. Predicting Performance

Due to the large number of potential predictors of post-test performance, it was necessary to determine which variables significantly predicted the outcome variable. Therefore, post-test percentage correct was regressed on gender, mode of help (visual/analytic), mental rotation composite score, math-fact retrieval composite score, and pre-test percentage correct. Controlling for all other variables, math-fact retrieval score, $b = 3.09, SE = 1.39, p = .03$, and pre-test score, $b = .46, SE = .10, p = .00$, were both significant predictors of post-test performance. Specifically, higher math-fact retrieval composite scores and pre-test scores were individually associated with higher

post-test scores. Combined, the independent variables accounted for 26% of the variance in post-test scores.

D. Help Seeking

Previous research suggested the presence of gender differences in certain measures of help seeking. A multivariate ANOVA was performed with number of total hints and number of mode-specific hints as dependent variables and gender and mode (visual/analytic) as independent variables. No significant effects for gender were found. However, there appeared to be a trend by which males ($M = 85.17$) asked for more help than their female counterparts ($M = 63.12$), $F(1, 127) = 1.54, p = .22$. Additionally, students in the visual condition ($M = 11.10$) saw more mode-specific hints than their counterparts in the analytic condition ($M = 7.11$), $F(1, 127) = 5.87, p = .02$. No significant interaction was found.

An additional ANOVA was performed with mode and gender as independent variables and several help seeking measures as the dependent variables. The dependent variables included the number of problems in which help was seen, the average number of hints viewed, the average time spent in these problems, and the average incorrect attempts made in problems where help was seen. There were also four variables which indicated when help was sought with respect to when an answer was given: number of problems where help was seen before a correct response, number of problems where help was seen before an attempt, number of problems where all help was seen before a correct response, and number of problems where all help was seen before any attempt. Table 6 summarizes these help seeking variables by gender and mode.

Table 6

Help Seeking Performance by Gender and Mode

Performance	Male		Female	
	Visual	Analytic	Visual	Analytic
Total Hints	89.55 (118.06)	80.80 (146.48)	67.44 (82.40)	63.44 (52.38)
Specific Hints	12.12 (12.43)	5.55 (6.03)	9.85 (9.58)	8.36 (7.08)
# Problems	15.23 (11.58)	11.30 (9.11)	15.54 (9.31)	12.61 (8.42)
Hints/Prob.	2.69 (.97)	3.03 (1.04)	2.64 (1.08)	2.79 (1.08)
Time/Help	85.53 (79.50)	64.11 (26.13)	86.41 (54.71)	99.59 (65.24)
Incorrect Attempts	.95 (.56)	.97 (.76)	1.18 (.74)	1.21 (.54)
Help Before Answer	14.55 (11.72)	10.60 (9.46)	14.33 (9.44)	11.50 (8.48)
Help Before Attempt	13.65 (11.63)	10.10 (8.84)	12.38 (9.25)	10.36 (8.26)
All Help/Answer	14.06 (11.68)	10.25 (9.36)	14.13 (9.38)	10.94 (8.34)
All Help/Attempt	12.87 (11.58)	9.55 (8.40)	11.77 (9.01)	9.28 (7.50)

Note. Standard deviations are in parentheses.

There was no significant main effect of gender nor were there any significant interactions. There was, however, a main effect of mode with respect to the number of problems where all help was seen before a correct answer was given. Specifically, those in the visual condition ($M = 13.90$) had a higher number of this type of problem than those in the analytic condition ($M = 10.75$), $F(1, 122) = 3.82, p = .05$. There were also several marginally significant variables: number of problems in which help was seen ($p = .06$), number of problems where help was seen before a correct answer ($p = .06$), and number of problems where all hints were seen before any attempt ($p = .09$). In each of these variables, there was a visible trend whereby those in the visual condition saw more help than those in the analytic condition.

Students were given a help seeking rating based on the total amount of hints viewed and the total number of mode-specific hints viewed. A median-split technique was used whereby students whose scores fell above the median were put in the “high help seeking group” and those whose scores fell below the median were placed in the “low help seeking group”. An ANOVA was performed with help seeking rating (low/high) as a between subjects variable and test performance (pre/post) as a repeated measure. There was a main effect of help rating, $F(1, 109) = 14.69, p = .00$, with students who asked for less help ($M = 4.16$) outperforming those who asked for a lot of help ($M = .31$) (see Figure 8). It should be noted that both groups, high and low help seeking, displayed substantial within group variability. There was no significant interaction.

It would be reasonable to expect that students who perform more poorly on the pre-test will ask for more help during the tutoring session. In order to test this

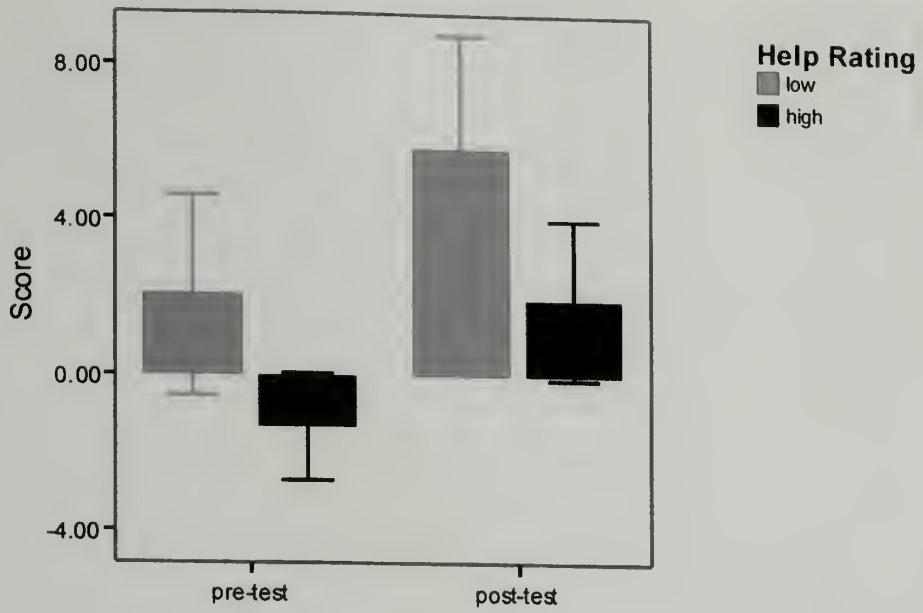


Figure 8. Pre- to post-test improvement by help rating.

hypothesis a multivariate ANOVA was performed with pre-test rating (high scorer/low scorer) as the between subjects independent variable and total number of hints requested and total number of mode specific hints viewed as the dependent variables. No significant effect was found.

CHAPTER IV

DISCUSSION AND CONCLUSIONS

A. Discussion

In-class observations of students working within the Wayang system were encouraging. Students appeared to enjoy the system, and a review of their answers to open-ended survey questions following completion of the experimental protocol strengthened this subjective observation. The majority of student comments were positive, though several suggestions were offered on ways to improve future versions of the system. For example, one student called the program “educational but fun” while several others commented on the usefulness of the help in breaking down problems. The main criticism of the system was the overall difficulty of the problems, both within the tutor and on the pre- and post-tests. The subjective student response to the system met experimenter expectations and the survey responses provided both praise and constructive criticism that will be carefully considered when designing future versions of the system.

Descriptive data recorded by the server provided a more objective look at student interaction within the Wayang system. Data for the pre- and post-tests revealed that the problems included on these tests were quite challenging for students. On average, students in the experimental group correctly answered only 23.16% of geometry problems and 22.96% of algebra problems on the pre-test. Similarly poor scores were found for the control group. Help seeking data, however, revealed that students actively sought help while in the tutoring module. Students in the visual group saw an average of 79.12 hints, while students in the analytic group saw 69.09

hints. Additionally, students, on average, made only one incorrect attempt before choosing the correct answer. Despite the difficulty of the problems presented, these results suggest that the majority of students did not click randomly until they found the correct answer. Instead it appears as though students requested help until they were able to come to the correct solution. The math-fact retrieval and mental rotation data revealed relatively low accuracy and a relatively long time spent working on each problem. Despite this poor performance, many students entered the modules repeatedly. Together these descriptive statistics give a comprehensive look at how students spent their time within the system, providing further evidence of student engagement.

The technical performance of the system was also satisfactory. One area of concern was the ability of the server to handle the load produced by approximately 30 students accessing the system simultaneously. However, despite technical difficulties in earlier versions of Wayang, the current system performed to expectations. There were no significant problems in any of the classrooms, and no reason to believe that the current results were influenced, either positively or negatively, by technical shortcomings.

1. Test Performance

The findings of the current study provide strong evidence for the veracity of the Wayang system as a tutoring tool for the SAT-Math exam. Students who worked with the system significantly improved their performance on tutored (geometry) problems, without significantly improving on control (algebra) questions. This within subject control removes the possibility that improvement seen in the experimental

group was due to extraneous factors such as the novelty of the tutoring environment, including the computerized tutor itself and the presence of researchers in the classroom. The benefit of the Wayang system was restricted to those questions that were explicitly tutored, emphasizing the system's effectiveness in tutoring specific skills.

As predicted, the control group did not improve on either geometry or algebra problems. These students were engaged in regular mathematics classroom activities during the week and took the pre- and post-test at the same time as students in the experimental group. The only difference between the two groups was that the control group did not experience the tutoring intervention that was presented to the experimental group. It was therefore concluded that the improvement seen in the experimental group was not simply due to general improvement over the school week or the effect of re-testing.

The Wayang tutor was effective in its goal of improving student performance on SAT-Math geometry problems. This finding demonstrates the potential benefit of adding tutoring systems to the regular classroom curriculum. Several previous studies in addition to the current study show that students can and do learn from these well-designed systems (Mayer et al., 2003; Moreno et al., 2001; Schofield, Eurich-Fulcer & Britt, 1994). However, students are not the only ones who will benefit from the addition of web-based tutors. Due to their easy-to-use nature and their proven effectiveness, teachers can utilize these tools in order to provide students with one-on-one help and increased control over their own learning, important tools teachers may have otherwise not been able to provide to students due to time or other constraints

(Schofield et al., 1994). Computerized tutors, and web-based tutors in particular with their flexible design, allow for many new and exciting learning opportunities for both teachers and students.

An interesting follow-up to this result concerns the way in which students' test profile changed from pre- to post-test. It was hypothesized that students in the experimental group would improve across all dimensions, increasing percent correct and decreasing percent incorrect and percent skipped. As predicted, students answered more questions correctly and skipped fewer questions from pre- to post-test. Interestingly, however, there was no significant change in the percentage of incorrect answers. One possibility is that students may have been attempting problems on the post-test that they would have otherwise skipped, and correctly answering them.

In order to explore this possibility, additional in-depth analyses of student performance are necessary. Specifically, the pre- and post-tests must be scored on a question-by-question basis. Correct, incorrect, and skipped problems could then be sorted on the basis of their component skills. A comparison of these lists from pre- to post-test could be used to determine if students were indeed attempting problems requiring certain skills which they had previously skipped in the pre-test. This, in effect, may help to reveal which skills the student was acquiring through his or her interaction with the tutor. Designers could then improve upon the help for specific skills which students are not effectively acquiring.

In addition to determining overall student improvement, a sub-goal of the current study was to explore potential gender differences in performance. It is a widely known finding that males and females achieve disparate scores on high-stakes

math achievement exams such as the SAT-Math exam, with males outperforming their female counterparts (Benbow, 1988; Hyde et al., 1990). The students in the current study were no exception, with males showing a marginally significant advantage over their female peers. This gender difference was seen in both the pre- and the post-test, and the difference did not diminish after use of the Wayang system. In this respect, the system fell short of its goal of decreasing the gender gap. However, the fact that all students, both males and females, showed general improvement from pre- to post-test is quite encouraging.

Further research is necessary to create educational materials that will enable females to achieve their full potential. Perhaps early intervention is the key to diminishing the gender effects on performance (McCormick & Wolf, 1993). Appropriate interventions with younger students are necessary to explore this possibility. Wayang's malleability with respect to core content opens up the potential for work with elementary school students.

Alternatively, it may be the case that with prolonged exposure to Wayang, females would show increased improvement relative to their male peers. Students in the current study were only exposed to the system for three class periods, providing a maximum of two hours spent working on the adventures and within the tutor. It is quite encouraging that in this short period of time students were able to significantly improve their performance. Additional time in the system may result in increased benefits and may perhaps even help to decrease the gender gap.

The current study also examined differences in learning styles. It is a well-known finding that students come into learning situations with a wide variety of

abilities and preferred learning styles (Casey et al., 1997; Gallagher, 1992). Wayang attempted to cater to students' needs by offering two distinct types of help: visual and analytic. The visual mode offered help based on strategies such as visual estimation while the analytic help proceeded in a more textbook-like manner. Due to the lack of a student model at the time of the current study, students were randomly assigned to one of the two conditions. Student performance within the Wayang system was not affected by the mode of help the student received, nor were there gender differences in response to help. However, it will be interesting to see if patterns emerge once students are assigned to groups based on their predetermined learning styles as opposed to relying on random assignment.

It should also be noted that the first few hints for both groups were the same. Hints only became differentiated after subsequent student requests. Students in the current study rarely asked for enough help to see these mode-specific hints. Therefore, it is possible that students of various abilities would respond differently given enough exposure to the specific types of help.

In addition to looking at gender and type of help, the current study aimed to determine what type of student, based on pre-test performance, was most likely to benefit from Wayang's instruction. Tests revealed that the tutor was most effective for those students who performed poorly on the pre-test. This is of great importance since tutors are often aimed at students who are most in need of help. Interestingly, the students who performed well on the pre-test did not significantly improve from pre- to post-test. It is unclear why these students did not improve, though there could be several reasons. Perhaps these students were not as motivated as their peers who

did not score as well on the pre-test. It may also be the case that the help was not as effective for this group. Future studies are necessary in order to weigh these possibilities and fully resolve the issue.

2. Mental Rotation and Math-Fact Retrieval

Numerous researchers have found that males and females differ on certain measures of spatial ability, including mental rotation (Linn & Peterson, 1985; Voyer et al., 1995). However, males and females in the current study did not differ on the online measure of mental rotation. Similarly, in the past, males have generally been found to outperform their female peers on measures of math-fact retrieval speed (Royer, Tronsky, Chan, et al., 1999). Contrary to previous results, there was a marginally significant result of females outperforming their male counterparts on the math-fact retrieval task.

Given the extensive prior research on both math-fact retrieval and mental rotation, it is difficult to resolve the current findings. It may be the case that the students in the current study differed in some way from those studied in previous research. Descriptive data for the math-fact retrieval task places average accuracy at 88.90% and average speed at 4.2 seconds per problem. Generally accuracy is found to be much higher, even in elementary school samples, often nearing 100% (Royer, Tronsky, Chan, et al., 1999). Additionally, the average time per problem appears quite slow for a high school sample. It may be the case that students in the current study were not focused on the task, thereby skewing the results. Future research with different samples will help to clarify this possibility.

Alternatively, with respect to math-fact retrieval, it could be that the online task used in this study was not tapping the same constructs as the conventional task. This is, in fact, a concern that was voiced before the current research was undertaken. Extraneous variables such as time taken to indicate a response with the mouse and task variables related to the use of a verification task as opposed to a production task are questionable.

Zbrodoff and Logan (1990) experimentally examined this issue. They conducted several experiments in order to explore their hypothesis that verification tasks are simply production tasks with an added comparison component. In production tasks participants are presented with a pair of digits (e.g., $4+5=$) and asked to produce their sum, usually verbally. The original math-fact retrieval task is of this variety. Verification tasks, however, require the participant to view a full equation (e.g., $4+5=8$) and indicate whether the equation is true or false. This was the type of task utilized in the current study. Zbrodoff and Logan set out to determine the exact relation between these two types of tasks. The researchers were unable to find any evidence that verification tasks are simply production task plus comparison. Alternatively, they found that participants use an entirely different strategy for this type of problem, opting to compare the entire equation against their memory. The authors concluded that verification tasks are rarely, if ever, simply production tasks with an added comparison component.

This finding has important implications for the current study. If verification tasks do not require production in its general form, which is the basis of math-fact retrieval, then the current results cannot be compared to previous results. It appears as

though the online version of the math-fact retrieval task is measuring something completely different than Royer, Tronsky, Chan, and colleagues' (1999) original math-fact retrieval task.

It is unclear whether the same problems plagued the mental rotation task. Unlike the math-fact retrieval task, the same requirements exist for the original version and the online version. Specifically, the student is responsible for determining if the figures can be mentally rotated to match one another. One potential problem is attentiveness to the task, which appeared to be an issue in the math-fact retrieval task. Barring this possibility, it is unlikely that extraneous task variables affected the mental rotation results. It could be the case that there truly were no gender differences in mental rotation, and, further, that performance on this task was not a good predictor of post-test performance. Research with additional samples is necessary to clarify this issue.

3. Predicting Performance

A major component of the second goal of the current study was to determine the factors which affect post-test score. Pre-test performance and math-fact retrieval composite score were the only significant predictors of post-test performance. Gender, help mode, and mental rotation composite score were all non-significant predictors. The correlation of pre- and post-test performance is not surprising. The finding that math-fact retrieval score is positively correlated to post-test performance is more interesting given the problems discussed previously. As was found by Royer, Tronsky, Chan, and colleagues (1999), students who were fast and accurate on the

math-fact retrieval task performed better on the post-test than those students who were slow and inaccurate.

This general finding is not surprising given the experimental literature (Royer, Tronsky, Chan, et al., 1999). What is surprising is the pattern of results described previously, with females outperforming their male peers. It is interesting that despite the variation in task demands, math-fact retrieval score is still found to significantly predict post-test performance when all other variables are held constant. It may be the case, as Royer, Tronsky, Chan, and colleagues (1999) found, that math-fact retrieval ability is a robust predictor of math test performance. Given Zbrodoff and Logan's (1990) findings, it may also be that males and females differ on verification tasks in addition to differing on production tasks, and perhaps these tasks can also provide insight on student performance.

4. Help Seeking

Differences were found in help seeking measures between the two help modes despite the fact that students in the visual condition did not outperform those in the analytic condition. Overall, as evidenced by various help seeking data, students in the visual condition were more engaged by the help than those in the analytic condition. Specifically, students in the visual condition saw more hints and interacted more with the help function than students in the analytic condition. Perhaps students were more attracted to the visual help since its strategies are not often presented in the classroom (Vincent, 2001). The analytic help would seem much more ordinary to students, as it is oftentimes the preferred mode of teaching. This finding, however, is tenuous at best considering that students did not see many mode-specific hints.

Contrary to previous findings, males and females in the current study did not differ in their help seeking behavior. There was no interaction between gender and type of help, indicating similar response to each help type regardless of gender. Again, it must be mentioned that students were not assigned to help groups by any specific guidelines nor did students become actively engaged in their specific mode of help. It will be interesting to see if any interaction occurs when these issues are resolved.

Due to the large amount of variability in help requests, students were divided into high and low help seeking groups in order to compare their pre- and post-test performance. Students who asked for less help performed better overall, though students who asked for a lot of help also showed significant improvement. One potential interpretation is that the high scoring students did not need the help the system was providing in order to attain a relatively high score. In the future, academic ability should be referenced in order to address this question.

B. Conclusions and Future Directions

The current study's main goal was to determine the effectiveness of the Wayang tutoring system with regards to improving SAT-Math performance in high school students. This goal was achieved, and Wayang was shown to be an effective tutoring system. Certain methodological issues may have affected more detailed results, including the role of math-fact retrieval and mental rotation ability in math test performance. These potential inconsistencies should be explored in future research.

With respect to math-fact retrieval and mental rotation, an experiment should be performed whereby the original tasks are compared with the online verification

versions. It is currently unclear whether the two tasks are correlated. The current research in addition to the work by Zbrodoff and Logan (1990) would indicate that in fact they are not. A systematic analysis could point to areas of disparity and perhaps improve the structure of current online systems.

Help mode differences may have been obscured due to the less than optimal approach of random assignment to conditions. In the future, intelligent components will be added to the Wayang tutoring system so that students may be assigned to help conditions based on their unique learning profile. This is an ongoing endeavor at the University of Massachusetts at Amherst. Data gathered from the current study is being utilized to inform future versions of the system. Specifically, difficulty ratings are being computed for each problem using average latencies and average percent correct. Student behavior, such as erratic clicking on answers or help, is being examined in order to define the student model. Finally, a set of rules will be created in order to determine which students, based on their previous performance in combination with in-system behavior, should see which problems and which type of help. Mental rotation and math-fact retrieval scores will also be used to inform a student model in the future. Students will be exposed to a learning environment tailored to their specific learning needs, allowing them, ideally, to reach their full potential.

The Wayang system could also be expanded to target younger students. The content could be created to target the needs of students in particular areas (e.g., MCAS preparation for sixth-grade students in Massachusetts), and of different ages. The

system was designed specifically to be amenable to these sorts of changes. This flexibility is a major advantage of web-based tutoring systems.

In summary, intelligent tutoring systems have a variety of potential uses that must be explored more fully. Preliminary studies such as the current research provide the necessary stepping stones for further investigation. The success of the current study in improving student performance attests to the utility of Wayang in particular and web-based tutoring systems in general. Perhaps with further research the full potential of computerized tutoring systems will be realized, to the benefit of students, teachers, and researchers.

REFERENCES

- Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20*, 327-343.
- Beal, C. (1999). Special issue on the math-fact retrieval hypothesis. *Contemporary Educational Psychology, 24*, 171-180.
- Benbow, C.P. (1988). Sex differences in mathematical reasoning ability among the intellectually talented: Their characteristics, consequences, and possible explanations. *Behavioral and Brain Sciences, 11*, 169-232.
- Bielinski, J., & Davison, M. (1998). Gender differences by item difficulty interactions in multiple-choice mathematics items. *American Educational Research Journal, 35*, 455-476.
- Casey, M., Nuttall, R., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology, 33*, 669-680.
- Casey, M., Nuttall, R., Pezaris, E., & Benbow, C. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology, 31*, 697-705.
- College Board. (2004). *How the test is scored*. Retrieved November 11, 2004, from <http://www.collegeboard.com/student/testing/sat/scores/understanding/howscored.html>
- Cooper, J., & Stone, J. (1996). Gender, computer-assisted learning, and anxiety: With a little help from a friend. *Journal of Educational Computing Research, 15*, 67-91.
- Cuban, L., Kirkpatrick, H., & Peck, C. (2001). High access and low use of technologies in high school classrooms: Explaining an apparent paradox. *American Educational Research Journal, 38*, 813-834.
- Eccles, J. (1997). User-friendly science and mathematics: Can it interest girls and minorities in breaking through the middle school wall? In D. Johnson et al. (Eds.), *Minorities and girls in school: Effects on achievement and performance* (Volume 1, pp. 65-104). Thousand Oaks, CA: Sage.
- Gallagher, A. (1992). *Sex differences in problem solving used by high scoring examinees on the SAT-M*. College Board Report No. 92-2, ETS RR No.92-33. New York: College Board Publications.

- Geary, D. (1999). Sex differences in mathematical abilities: Commentary on the math-fact retrieval hypothesis. *Contemporary Educational Psychology, 24*, 267-274.
- Green, S., & Wolf, I. (1998). *Barron's SAT I How to Prepare for the SAT I 20th Edition*. Hauppauge, NY: Barron's Educational Series, Inc.
- Hyde, J. (1981). How large are cognitive gender differences? A meta-analysis using w and d. *American Psychologist, 36*, 892-901.
- Hyde, J., Fennema, E., and Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139-155.
- Kimball, M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105*, 198-214.
- Linn, M., & Peterson, A. (1985). Emergence of characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479-1498.
- Manger, T., & Eikeland, O. (1998). The effects of spatial visualization and students' sex on mathematical achievement. *British Journal of Psychology, 89*, 17-25.
- Mayer, R., Sobko, K., & Mauton, P. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology, 95*, 419-425.
- McCormick, M., & Wolf, J. (1993). Intervention programs for gifted girls. *Roeper Review, 16*, 85-88.
- Moreno, R., Mayer, R., Spires, H., & Lester, J. (2001). The case of social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19*, 177-213.
- Princeton Review (2000). *The Princeton Review Course Manual & Workbook for SAT Math*. Princeton Review Management, L.L.C.
- Reuhkala, M. (2001). Mathematical skills in ninth-graders: Relationship with visuo-spatial abilities and working memory. *Educational Psychology, 21*, 387-399.
- Rosser, P. (1989). *The SAT gender gap: Identifying the causes*. Washington, DC: Center for Women's Policy Studies.
- Royer, J., Tronsky, L., Chan, Y., Jackson, S., & Marchant, H. (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology, 24*, 181-266.

- Royer, J., Tronsky, L., Marchant, H., & Jackson, S. (1999). Reply to the commentaries on the math-fact retrieval hypothesis. *Contemporary Educational Psychology*, 24, 286-300.
- Schofield, J., Eurich-Fulcer, R., & Britt, C. (1994). Teachers, computer tutors, and teaching: The artificially intelligent tutor as an agent for classroom change. *American Educational Research Journal*, 31, 579-607.
- Shea, D., Lubinski, D., & Benbow, C. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, 93, 604-614.
- Stanley, J., & Benbow, C. (1983). 15:1 certainly isn't "catching up". *Psychological Reports*, 52, 656.
- Subrahmanyam, K., & Greenfield, P. (1994). Effect of video game practice on spatial skills in boys and girls. *Journal of Applied Developmental Psychology*, 15, 13-32.
- Vandenberg, S. (1978). *Vandenberg Test of Mental Rotation*. (Unpublished test)
- Vincent, J. (2001). The role of visually rich technology in facilitating children's writing. *Journal of Computer Assisted Learning*, 17, 242-250.
- Voyer, D., Voyer, S., & Bryden, M. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117, 250-270.
- Webb, R.M., Lubinski, D., & Benbow, C.P. (2002). Mathematically facile adolescents with math-science aspirations: New perspectives on their educational and vocational development. *Journal of Educational Psychology*, 94, 785-794.
- Weiner, N., & Robinson, S. (1986). Cognitive abilities, personality and gender differences in math achievement of gifted adolescents. *Gifted Child Quarterly*, 30, 83-87.
- Wigfield, A., & Byrnes, J. (1999). Does math-fact retrieval explain sex differences in mathematical test performance? A commentary. *Contemporary Educational Psychology*, 24, 275-285.
- Willingham, W., & Cole, N. (Eds.). (1997). *Gender and fair assessment*. Mahwah, NJ: Erlbaum.

Zbrodoff, N., & Logan, G. (1990). On the relation between production and verification tasks in the psychology of simple arithmetic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16, 83-97.

