Improving Methods of Teaching Mathematics in Middle School

An Interactive Qualifying Project

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

in

Computer Science

by

_____

Abraao Lourenco

April 2006

APPROVED:

_____

Professor Neil Heffernan, IQP Advisor

Abstract

This Interactive Qualifying Project discusses two different methods for improving math learning in the classroom. The first method is a scaffolding strategy emphasizing active participation. The second method is solving equations by algebraic manipulation. The experiments indicate that employing the scaffolding strategy can improve students' performance in math.

Table of Contents

Table of Figures

Introduction

Beginning in 2003 successful completion of the Mathematics as well as the English portions of the Massachusetts Comprehensive Assessment system exam (i.e. the MCAS) is required before a student graduates. Although policy makers have sold MCAS "as a necessary 'stick' to get students to work hard and take school seriously, MCAS is as likely to drive them away as to motivate them to work harder" [Crisis].

Many students who failed the MCAS are afraid of repeated humiliating failures [Crisis]. A 2002 report from Worcester's Center for Community Performance Measure showed that many of the students who failed scored very close to a passing grade: 59% of the students who took the math exam scored 4 to 2 points lower than the necessary passing grade [Worcester].

The United States has been the world's leader in research and development, but according to the Task Force on the Future of American Innovation, the USA may lose its leadership role in science and innovation if something is not done soon [Innovation]. One of the six benchmarks identified by the task force was the proportion of US citizens graduating in Science & Engineering; it has declined by 10% between 1994 and 2001 [Innovation]. In order to support innovation in the future, the task force recommended increasing the budget of key agencies such as the National Science Foundation by 10%-12%.

The Department of Education has given a $1.4 million federal grant to Prof. Neil Heffernan from WPI, Kenneth R. Koedinger and Brian Junker from Carnegie Mellon University, and Steven Ritter from Carnegie Learning, Inc., to test a system that simultaneously assesses and assists. This system is called the "ASSISTment Project", as it allows students to get instructional **assist**ance on MCAS items while at the same time

provide **assessment** of students to teachers, and thus the term "ASSISTment" [ASTMain]. This system also allows the execution of experiments that evaluate different teaching strategies.

This paper analyzes students' performance in two experiments. In the first experiment students were presented with a number of questions using one of two strategies, scaffolding based or hint based. Student performance was measured by creating transfer questions that were similar to the experimental questions but used the scaffolding strategy. Similarly, the questions in the second experiment employed either an algebraic manipulation strategy or a variable substitution strategy. Student performance in the second experiment was also measured using transfer questions. The goal of these experiments was to determine the teaching method that yielded the best student performance. An analysis of the student performance data collected indicates that scaffolding strategy yields the best student performance in the first experiment, as does the algebraic manipulation strategy for the second experiment. The analysis of the collected student performance data and the conclusion of the experiments are presented in this paper.

Background

The first experiment in this paper investigated scaffolding as a means to improve student performance in math. Vygostky explained how learning can be facilitated with his concept of "zone of proximal development". The "zone of proximal" development for a particular piece of information is that time in the learner's life when she/he is ready to learn a particular piece of information but does not have all the prerequisites or other information that is needed to acquire the information without assistance [Scaffolding1].

Vygostky asserts that the teacher or facilitator can provide this information by helping the learner build a structure into which to put the new information; the act of building this structure is called scaffolding [Scaffolding1]. Put more simply, "zone of proximal scaffolding is an interactive process by which a learner is assisted by others (teachers or peers) to acquire a skill which cannot be acquired without assistance". Scaffolding can be provided by teachers, peers, computer screens, etc. In this experiment scaffolding was provided by the ASSISTment system.

According to Bull et al scaffolding facilitates learning in the following ways: it reduces ambiguity and helps the learner connect the new information to what they already know. By reducing learning ambiguity and helping the learner place the new information in their own internal representation of the knowledge, scaffolding aids in the construction of meaningful, structured relationships between what the learner knows and the new knowledge, transforming the new information into personal knowledge [Scaffolding1].

Scaffolding can be provided in several different ways. Scaffolding may provide explanations of certain core concepts of the question when the learner does not understand the question, or contributes ideas and suggestions on what approach the

learner should take; the ASSISTment system supports this in the form of hints. Some sophisticated scaffolding systems are able to answer questions by the learner; this is the role usually taken by the teacher when using the ASSISTment system, as it currently does not handle natural language parsing for student questions. Scaffolding can also be interactive and invite the learner to think about how to think through the problem and show what ideas and concepts are important for a particular question, helping those who may know some concepts but lack a key fact to solve the question; this is the condition referred to as scaffolding in the first experiment. In this role scaffolding provides evidence that the learner is following the right path, or statements refuting the answer provided if it is incorrect.

The second experiment investigated algebraic manipulation as a tool for teaching students to understand how to solve equations. Algebraic manipulation is a means whereby a student uses the rules of arithmetic to manipulate the variables and numbers in an equation to make it easier to understand and to ultimately solve it. Variable substitution on the other hand, involves plugging numbers into the variables and evaluating both sides of the equation to check for their validity. While substituting numbers in for variables yields an answer eventually (assuming the student is given a finite number of choices to try, at least one of which is correct), it doesn't give all the answers and it does not develop the necessary skills for understanding algebraic equations. Algebraic manipulation, on the other hand, develops skills that can be used in new situations to solve different equations. There is no literature on the subject of algebraic manipulation in comparison to substituting in numbers as it is "common belief" that algebraic manipulation is the most effect way to teach students. We hope to provide some empirical evidence on the subject.

Metrics and Scenarios

The questions for each experiment were presented using the ASSISTment system. The ASSISTment system is a web based intelligent tutoring system that assesses and assists students simultaneously. Students at several Worcester middle schools used the system for one hour every week. During that hour students worked on the curriculums assigned to them.

All the questions were organized into two different curriculums. The curriculums were organized so that all the questions in the curriculum shared the same topic. The algebraic manipulation and variable substitution questions were on the subject of number lines, using inequalities and equations to represent number lines. The scaffolding vs hints curriculum the problems involved word problems and simple arithmetic. By doing this we gave the students several opportunities to learn the concept, exposing them to the same type of problem several times. This was necessary in order to evaluate the student's learning at the end.

The purpose of each curriculum was to determine the effectiveness of two different learning strategies. This is done by dividing the curriculum in a pre-test section, two experimental sections and one transfer section. The pre-test section was used to gauge the student's knowledge of the material and consisted of very basic problems. The purpose of the experimental sections was to teach students a specific strategy for approaching a problem so that we could later evaluate their performance on similar questions and thus judge the validity of the approach they learned in the experimental section. That was the purpose of the transfer section.

The transfer section was perhaps the most important part of a curriculum. It presented new material (about five or six items) which invited the student to use the

techniques employed in the specific section in the experimental section on this new material. By analyzing which group of students (each group assigned to a different experimental subsection) improved the most in the transfer section compared to the pre-test section, it was possible to determine which teaching strategy improved students' performance. For example, if the students who used algebraic manipulation did better on the problems in the transfer section than the student who used variable substitution then that would imply a correlation between student performance and learning algebraic manipulation.
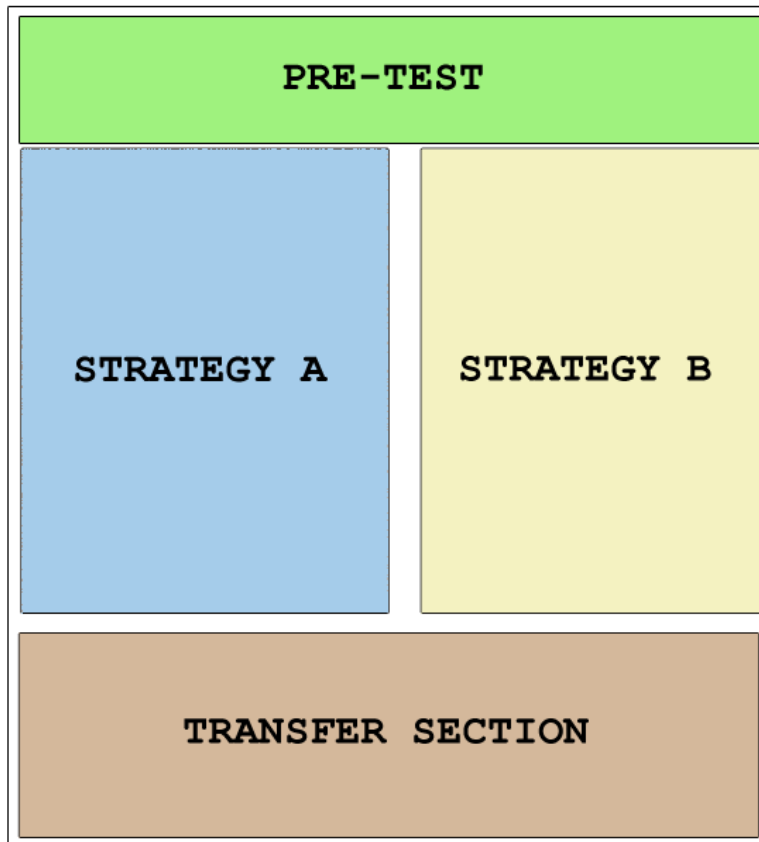
Figure 1 - A conceptual diagram of a curriculum

In the scaffolding strategy experiment, one of these sections had questions that used a scaffolding strategy, and the other section had questions that used a hint strategy. Similarly, in the algebraic manipulation experiment one of the section contained questions that encouraged a variable substitution approach while the other section suggested algebraic manipulation. Each of these subsections had about five or six problems.

Students who received the experimental section with the scaffolding strategy were presented with a series of questions that broke down the problem into smaller steps, each one dealing with a more specific concept. This only happened if they got the original question wrong or if they asked for a hint in the original question. An example is shown in figure 2. Similarly, students who were assigned the hint strategy section were given a number of hints that explained the steps necessary to arrive at the answer without asking the students any questions. An example is shown in figure 3. The problems in the transfer section for this experiment used the scaffolding strategy.

$$y = \frac{100}{x} + 50$$

In the equation shown above, **x** represents a positive real number. As the value of **x** gets larger, what happens to the value of **y**?

- ○ A. The value of y stays the same.
- ○ B. The value of y increases.
- ○ C. The value of y approaches 50.
- ● D. The value of y approaches 100.

Hmm, no.

Let me break this down for you.

_____

$$\frac{100}{x}$$

What happens to the value of the expression above when the value of **x increases**?

- ○ It decreases.
- ○ It increases.

[Submit]

[Done] [Hint] [More]

Figure 2 - An item following the scaffolding strategy

$y = \frac{100}{x} + 50$

In the equation shown above, **x** represents a positive real number. As the value of **x gets larger**, what happens to the value of **y**?

- ○ A. The value of y stays the same.
- ○ B. The value of y increases.
- ○ C. The value of y approaches 50.
- ● D. The value of y approaches 100.

[Submit]

Sorry, that is not correct.

Think about what happens to the value of the expression above when the value of x increases.

[Done] [Hint] [More]

Figure 3 - A problem that uses the hint based strategy

In the algebra versus substitution experiment students who were assigned the algebraic manipulation strategy received hints and questions that asked them to manipulate inequalities and equations in order to arrive at the answer. They were asked questions such as "What number do you need to use to divide both sides of the equation $3x = 15$ in order to isolate x?". On other hand students who were assigned the variable substitution experimental section were given some numbers as choices and were asked to plug them into the equations and inequalities in order to arrive at the answer. They were asked questions such as "What does the left side of the equation $3x = 15$ look like if we plug in 3 for x?". The problems in the transfer section for this experiment used the algebraic manipulation strategy.

At the start of each class students would login to the ASSISTment website. Each student was associated with a specific class led by a teacher, and each teacher assigned one of the experimental curriculums to the class. Once logged in, each student in the class was prompted to start the curriculum. When a student started the curriculum he/she was assigned a random subsection in the experimental section. This was done in order to minimize the chance that specific group would consist specifically of students who performed well with the given strategy, or that a specific group consisted only of unmotivated students. After completing the experimental section, both groups of students then did the items in the transfer section.

The ASSISTment system collects every action the student inputs into the system. This includes incorrect answers, the number of hints requested, the time taken to complete each question and the curriculum as a whole. We can then analyze the performance of the student or the class on the curriculum and assess which teaching strategy was more effective for a given experiment. Although data was collected for

every student who logged into the system, only data from students who completed the

entire curriculum was considered.

	ACTION_ID	ACTION_TIME	SCHOOL_ID	TUTOR_ID	CLASS_ID	USER_ID	ACTION_INITIATO
1	137320	10-MAY-04 08.19.59.000000 AM	0101002		Period 1		
2	137319	10-MAY-04 08.19.59.000000 AM	0101002		Period 1		77
3	137318	10-MAY-04 08.19.59.000000 AM	0101002		Period 1	2	132
4	137317	10-MAY-04 08.19.59.000000 AM	0101002		Period 1	2	132
5	137316	10-MAY-04 08.19.58.000000 AM	0101002		Period 1		75
6	137315	10-MAY-04 08.19.58.000000 AM	0101002		Period 1		75
7	137314	10-MAY-04 08.19.57.000000 AM	0101002		Period 1		76
8	137313	10-MAY-04 08.19.57.000000 AM	0101002		Period 1		76
9	137312	10-MAY-04 08.19.55.000000 AM	0101002		Period 1		169
10	137311	10-MAY-04 08.19.55.000000 AM	0101002		Period 1		169
11	137310	10-MAY-04 08.19.54.000000 AM	0101002		Period 1		75
12	137309	10-MAY-04 08.19.54.000000 AM	0101002		Period 1		75
13	137308	10-MAY-04 08.19.52.000000 AM	0101002		Period 1		7
14	137307	10-MAY-04 08.19.51.000000 AM	0101002		Period 1		6
15	137306	10-MAY-04 08.19.48.000000 AM	0101002		Period 1		
16	137305	10-MAY-04 08.19.48.000000 AM	0101002		Period 1		
17	137304	10-MAY-04 08.19.48.000000 AM	0101002		Period 1		188
18	137303	10-MAY-04 08.19.47.000000 AM	0101002		Period 1		107
19	137302	10-MAY-04 08.19.47.000000 AM	0101002		Period 1		107
20	137301	10-MAY-04 08.19.47.000000 AM	0101002		Period 1		187
21	137300	10-MAY-04 08.19.47.000000 AM	0101002		Period 1		187
22	137299	10-MAY-04 08.19.45.000000 AM	0101002		Period 1		106
23	137298	10-MAY-04 08.19.45.000000 AM	0101002		Period 1		106

Figure 4 - A sample of the raw data collected by the ASSISTment system

Results

Factors Affecting Results

Although students were randomly assigned to experimental sections in order to eliminate bias in the results, there are still some factors that might have affected the final conclusion.

For various reasons, such as lack of motivation or knowledge, some students were not able to complete one or both of the curriculums containing the experiments. If more students finished from a specific experimental section than from the other experimental section finished the curriculum this can cause a selection effect. This causes the final result to be biased since there weren't as many data points for both experimental conditions. To deal with this, we randomly selected a proportional number of results for each experimental condition in an experiment.

Some students perform better than others. This is hardly something that should be tried to be contained, but it still has the possibility to skew the results. This was mitigated by randomly assigning each student to an experimental section.

Not every student was motivated to complete the curriculums. This either caused the students to leave the curriculum unfinished as mentioned above regarding the selection effect, or led the student to "game" the system as mentioned in the next paragraph. In the first case, the student data was simply not considered, and the latter case there was not much that could be done with regards of throwing out the data; the effect was hopefully minimized by the random assignment of students to experimental sections.

One of the big problems was student "gaming". Gaming can be described as the student taking advantage of the technical setup of the system in order to improve their performance. It is more likely however, that those students who were gaming were doing

so out of boredom and a desired to finish the assignment as quickly as possible with a minimum of effort. Examples of gaming include requesting hints repeatedly until the "bottom out" hint is given. Usually every question has a bottom out hint that gives the desired answer for a question; this is done so that students don't get stuck. This gaming method is usually curtailed by alert teachers who monitor the number of hints requested by students through the reporting interface. Another gaming method was to force the system to restart a problem by reloading the curriculum listing screen. This method was quickly fixed by the WPI team using a replay system which "replays" a student's actions if the question is accessed again within a given timeframe. More recently another system has been implemented for detecting student gaming that doesn't require teacher intervention. When the system detects that a student is requesting hints too quickly only in order to get through the curriculum, it pops up a message saying that the student should be making a better effort to learn from the hints.

Some of the data we collected represented students who had not completed every question. Thus, we removed from the results students who did not have a complete transfer test, or who were missing a single question, indicating a software bug. We have no reason to think this software bug was disproportional by conditional.

Another issue is balance between students in each experimental condition. We expected both groups of students to perform similarly in terms of pre-test score. Any imbalance between groups in the pretest score would cause the calculated gain to be skewed in favor of the group with the lower pre-test score.

Data Analysis

In order to analyze the data collected we first processed the actions collected by the ASSISTment system and aggregated the data by student and question. That is, for each student / question pair we determined whether the student answered the question correctly on the first try and how long they took to solve the question.

After processing the raw data we created a pivot table that contained the student id as a row, the problem name and condition as a column, and the result (whether the student answered the question correctly) as the data. This enabled us to easily identify students who hadn't finished the curriculum or who for some reason were missing data; these students were removed from the results.

Once we had the pivot table, we calculated the gain for each student. The gain is calculated by subtracting the total number of questions answered correctly in the pretest section from the total number of questions answered correctly in the transfer section. The data was then imported into a statistical analysis tool called StatView and analyzed as described in the next sections.

Hints vs Scaffolding

Performance Measure

An unpaired t-test can be used to test the null hypothesis that the means for two populations is equal. That is, a t-test can tell us if the variation between two groups is significant. In this experiment we used a t-test to see if the variation in gain between the group who did the experimental section using hints and the group who did the section using scaffolding was statistically significant.

Unpaired t-test for Gain
Grouping Variable: Condition
Hypothesized Difference = 0

	Mean Diff.	DF	t-Value	P-Value
h, s	-.447	145	-1.932	.0553

Group Info for Gain
Grouping Variable: Condition

	Count	Mean	Variance	Std. Dev.	Std. Err
h	82	1.768	2.057	1.434	.158
s	65	2.215	1.797	1.340	.166

Figure 5 - Unpaired t-test for Gain (Hints vs Scaffolding)

The so called "null" hypothesis is that there is no difference in the mean gains scores; that is, the samples for both groups were drawn from the same underlying normal distribution. But the differences in the means of 1.768 and 2.215 seems large, so we performed a t-test to determine the chance that such a large difference in means is purely coincidental. The p-value for the t-test was 0.0553; that means that there is a 5.5% chance that if you assume these means were drawn from the same underlying distribution that you get the difference in means. This value is close enough to 0.05, the standard cut of to use the accepted term "statistically significant" that we will accepted this is as a statistically significant result. We also calculated the effect size, which is equal to the difference between the means (2.215 - 1.768 = 0.445) normalized by the standard deviation of the control group, the students who did the experimental section using the hints strategy (0.445 / 1.4 = 0.318) which is generally interpreted to mean a small effect size. A 95% confidence interval on the effect size is -0.01 to 0.65, indicating that the effect size of 0.31 is somewhere within the range of 0.00 to 0.62. The fact that this range includes zero means that we do not believe the effect size is greater than 0 with 95%

confidence. This corresponds to our p-value of 0.0553, which is just at the beginning of being statistically significant.

We also used a one-way ANOVA to test whether the experimental condition significantly affected the gain. The ANOVA table contains the degrees of freedom (i.e. sample size – 1), one F-Value and one P-Value. The P-Value tells us whether student performance depends on the experimental condition (hints vs scaffolding). The P-Value of 0.0553 tells us that student performance is statistically different between each experimental condition.

The second table has some statistical data for each condition such as the number of student responses for each experimental condition, mean value for the gain, standard deviation, and standard error. The higher mean for the scaffolding experimental condition indicates that on average, students who did the questions with the scaffolding strategy had a higher gain than other students. The standard deviation shows us that the variability in each condition in not very large. The standard error is a measure of the size of the variation in the sample statistic over all samples of the same size as the study sample. The values for standard error are small; this is expected as we only have two samples.

ANOVA Table for Gain

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	7.248	7.248	3.732	.0553	3.732	.469
Residual	145	281.582	1.942				

Means Table for Gain
Effect: Condition

	Count	Mean	Std. Dev.	Std. Err.
h	82	1.768	1.434	.158
s	65	2.215	1.340	.166

Interaction Bar Plot for Gain
Effect: Condition
Error Bars: 95% Confidence Interval
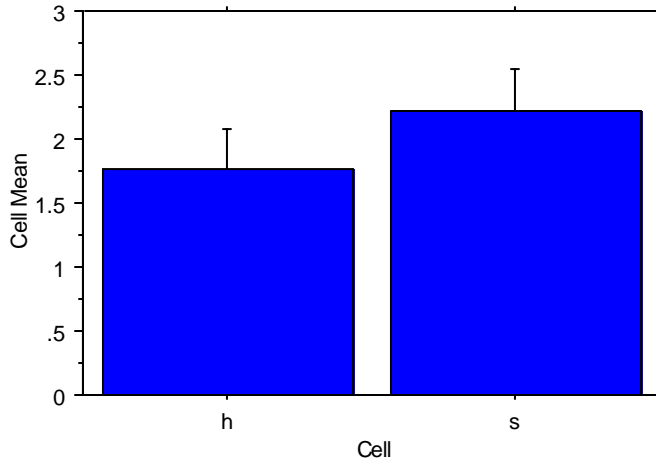
Fisher's PLSD for Gain
Effect: Condition
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value
h, s	-.447	.457	.0553

Figure 6 - ANOVA table for Gain (Hints vs Scaffolding)

Time Measure

We also analyzed the difference in the total time taken for a student to do all the problems in a given experimental section. In this experiment we used a t-test to see if the variation in total time taken between the group who did the experimental section using

hints and the group who did the section using scaffolding was statistically significant. The null hypothesis is that there is no difference in the mean time; that is, the samples for both groups were drawn from the same underlying normal distribution. The p-value for the t-test was 0.0020; that means that there is a 0.2% chance that if you assume these means were drawn from the same underlying distribution that you get the difference in means. The data shows that the average time taken by students who did the problems in the experimental section using the hints strategy is 481.139 seconds, while the average time taken by students who did the problems in the experimental section using the scaffolding strategy is 619.535 seconds.

Unpaired t-test for Total time
Grouping Variable: Condition
Hypothesized Difference = 0

	Mean Diff.	DF	t-Value	P-Value
Hints, Scaffolding	-138.396	263	-3.125	.0020

Group Info for Total time
Grouping Variable: Condition

	Count	Mean	Variance	Std. Dev.	Std. Err
Hints	151	481.139	89919.667	299.866	24.403
Scaffolding	114	619.535	177169.915	420.916	39.422

Figure 7 - t-test for total time (Hints vs Scaffolding)

We also used a one-way ANOVA to test whether the experimental condition significantly affected the total time taken. The ANOVA table contains the degrees of freedom (i.e. sample size – 1), one F-Value and one P-Value. The P-Value tells us whether student performance depends on the experimental condition (hints vs scaffolding). The P-Value of 0.0020 tells us that student performance is statistically different between each experimental condition.

ANOVA Table for Total time

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	1244179.651	1244179.651	9.765	.0020	9.765	.895
Residual	263	33508150.439	127407.416				

Means Table for Total time
Effect: Condition

	Count	Mean	Std. Dev.	Std. Err.
Hints	151	481.139	299.866	24.403
Scaffolding	114	619.535	420.916	39.422

Interaction Bar Plot for Total time
Effect: Condition
Error Bars: 95% Confidence Interval
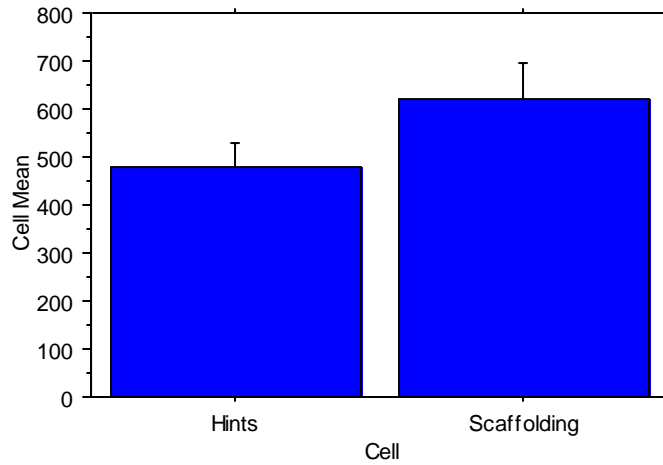
Fisher's PLSD for Total time
Effect: Condition
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
Hints, Scaffolding	-138.396	87.203	.0020	S

Figure 8 - ANOVA table for total time (Hints vs Scaffolding)

Algebraic Manipulation vs Variable Substitution

Performance Measure

In this experiment we used a t-test to see if the variation in gain between the group who did the experimental section using algebra and the group who did the section using substitution was statistically significant.

Unpaired t-test for Gain
Grouping Variable: Condition
Hypothesized Difference = 0

	Mean Diff.	DF	t-Value	P-Value
a, s	.473	159	1.983	.0490

Group Info for Gain
Grouping Variable: Condition

	Count	Mean	Variance	Std. Dev.	Std. Err
a	82	.878	2.207	1.486	.164
s	79	.405	2.372	1.540	.173

Figure 9 - Unpaired t-test for Gain (Algebra vs Substitution)

The so called "null" hypothesis is that there is no difference in the mean gains scores; that is, the samples for both groups were drawn from the same underlying normal distribution. But the differences in the means of 0.878 and 0.405 seems large, so we performed a t-test to determine the chance that such a large difference in means is purely coincidental. The p-value for the t-test was 0.0490; that means that there is a 4.9% chance that if you assume these means were drawn from the same underlying distribution that you get the difference in means. We also calculated the effect size, which is equal to the difference between the means (0.878 – 0.405 = 0.473) normalized by the standard deviation of the control group, the students who did the experimental section with the substitution strategy (0.473 / 1.540 = 0.307) which is generally interpreted to mean a small effect size. A 95% confidence interval on the effect size is 0.00 to 0.62, indicating

that the effect size of 0.31 is somewhere within the range of 0.00 to 0.62. The fact that this range includes zero means that we do not believe the effect size is greater than 0 with 95% confidence. This corresponds to our p-value of 0.0553, which is just at the beginning of being statistically significant.

We also used a one-way ANOVA to test whether the experimental condition significantly affected the gain. The ANOVA table contains the degrees of freedom (i.e. sample size – 1), one F-Value and one P-Value. The P-Value tells us whether student performance depends on the experimental condition (algebra vs substitution). The P-Value of 0.0490 tells us that student performance is statistically different between each experimental condition.

The second table has some statistical data for each condition such as the number of student responses for each experimental condition, mean value for the gain, standard deviation, and standard error. The higher mean for the algebra experimental condition indicates that on average, students who did the questions with the algebra strategy had a higher gain than other students. The standard deviation shows us that the variability in each condition in not very large. The standard error is a measure of the size of the variation in the sample statistic over all samples of the same size as the study sample. The values for standard error are small; this is expected as we only have two samples.

ANOVA Table for Gain

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	9.001	9.001	3.934	.0490	3.934	.491
Residual	159	363.818	2.288				

Means Table for Gain
Effect: Condition

	Count	Mean	Std. Dev.	Std. Err.
a	82	.878	1.486	.164
s	79	.405	1.540	.173

Interaction Bar Plot for Gain
Effect: Condition
Error Bars: 95% Confidence Interval
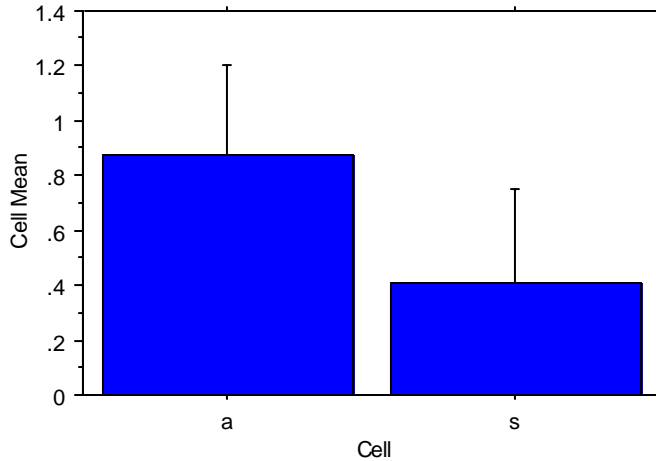
Fisher's PLSD for Gain
Effect: Condition
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
a, s	.473	.471	.0490	S

Figure 10 - ANOVA table for Gain (Algebra vs Substitution)

Time Measure

We also analyzed the difference in the total time taken for a student to do all the
problems in a given experimental section. In this experiment we used a t-test to see if the
variation in total time taken between the group who did the experimental section using

algebra and the group who did the section using substitution was statistically significant. The null hypothesis is that there is no difference in the mean time; that is, the samples for both groups were drawn from the same underlying normal distribution. The p-value for the t-test was 0.7336; that means that there is a 73.36% chance that if you assume these means were drawn from the same underlying distribution that you get the difference in means. This means that the difference in total time taken by the two groups is not statistically significant.

Unpaired t-test for Total time
Grouping Variable: Condition
Hypothesized Difference = 0

	Mean Diff.	DF	t-Value	P-Value
Algebra, Substitution	78.106	36	.343	.7336

Group Info for Total time
Grouping Variable: Condition

	Count	Mean	Variance	Std. Dev.	Std. Err
Algebra	17	945.059	579358.434	761.156	184.607
Substitution	21	866.952	413146.848	642.765	140.263

Figure 11 - t-test for total time (Algebra vs Substitution)

We also used a one-way ANOVA to test whether the experimental condition significantly affected the total time taken. The ANOVA table contains the degrees of freedom (i.e. sample size – 1), one F-Value and one P-Value. The P-Value tells us whether the total time taken by students on an experimental section depends on the experimental condition (algebra vs substitution). The P-Value of 0.7336 tells us that student performance is not statistically different between each experimental condition.

ANOVA Table for Total time

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	57313.685	57313.685	.118	.7336	.118	.063
Residual	36	17532671.894	487018.664				

Means Table for Total time
Effect: Condition

	Count	Mean	Std. Dev.	Std. Err.
Algebra	17	945.059	761.156	184.607
Substitution	21	866.952	642.765	140.263

Interaction Bar Plot for Total time
Effect: Condition
Error Bars: 95% Confidence Interval
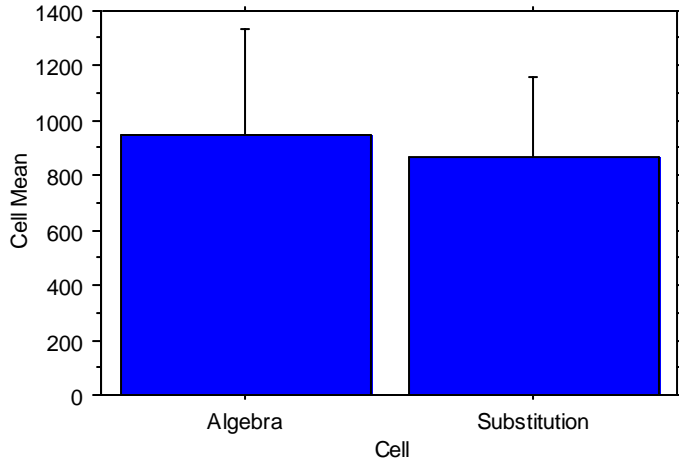
Fisher's PLSD for Total time
Effect: Condition
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value
Algebra, Substitution	78.106	461.762	.7336

Figure 12 - ANOVA table for total time (Algebra vs Substitution)

Pre-Test Balance

Although an analysis of the gain indicates that there is a statistically significant difference in gain between the two experimental groups, an analysis of the pre-test shows that this might not be real. The ANOVA table in figure 13 shows an analysis of the total

score for the transfer section items dealing with equation solving based on pretest scores.

As the P-values show, the conditions don't seem to statistically significant.

ANOVA Table for TransferEquationSovleing

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	.034	.034	.063	.8014	.063	.057
Pretest Total	1	6.225	6.225	11.568	.0009	11.568	.941
Condition * Pretest Total	1	.274	.274	.509	.4768	.509	.106
Residual	157	84.483	.538				

Means Table for TransferEquationSovleing
Effect: Condition

	Count	Mean	Std. Dev.	Std. Err.
a	82	1.305	.715	.079
s	79	1.278	.800	.090

Interaction Bar Plot for TransferEquationSovleing
Effect: Condition
Error Bars: 95% Confidence Interval
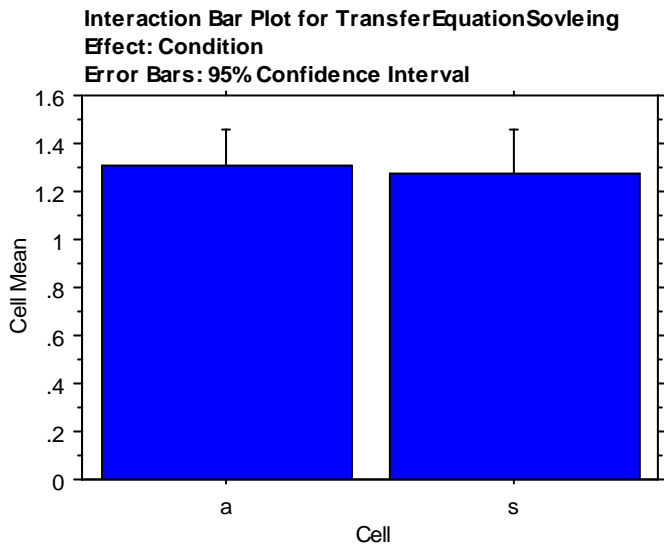
Fisher's PLSD for TransferEquationSovleing
Effect: Condition
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value
a, s	.026	.228	.8197

Figure 13 - ANOVA table for pretest analysis

Conclusion

In this paper we attempted to produce some evidence determining the effectiveness of two pairs of different learning methods. The first pair was scaffolding and hint based assistance, and the second pair was algebraic manipulation and variable substitution. For each pair, we ran an experiment where a class of students got a curriculum composed of a pre-test section, two experimental sections, each with one condition of the pair, and a transfer section. Half of the class of students did the questions in one experimental section of the curriculum, and the other half did the other experimental section. Then the whole class did the transfer section. The students' performance on each item was used to evaluate the merit of each teaching strategy.

The first experiment involved comparing scaffolding against hints. Background research indicated that scaffolding is a more effective teaching strategy than providing only hints to the students. The second experiment compared algebraic manipulation against variable substitution. According to "common belief", it is more effective to teach students how to solve questions, that is, teach them how to do algebraic manipulation, than to tell them to plug in number at random. While the later strategy might be effective when given a small number of choices, it becomes problematic if that is not the case.

This research shows that scaffolding is a more effective teaching strategy than just using hints, as well as that solving questions using algebra is better than just guessing the answer. While there were certain issues that could have affected the results, these were mitigated as possible. The results showed in this experiment can be used by educators to formulate curriculum with effective learning strategies.

.

References

[Algebra] Kirshner, David. <u>The Structural Algebra Option: a Discussion Paper</u>. Annual Meeting of the American Educational Reserach Associaton, Apr. 1993, American Educational Research Association. 26 Aug. 2005 <http://www.eric.ed.gov>.

[ASTMain] Razzaq, L., M. Feng, G. Nuzzo-Jones, N. T. Heffernan, K. T. Koedinger, B. Junker, S. Ritter, A. Knight, C. Aniszczyk, S. Choksey, T. Livak, E. Mercado, T. E. Turner, R. Upalekar, J. A. Walonoski, M. A. Macasek, and K. P. Rasmussen. <u>The Assistment Project: Blending Assessment and Assisting</u>. Proceedings of the 12th Artificial Intelligence in Education, 2005, Worcester Polytechnic Institute. <http://www.assistment.org/portal/project/papers/AIED2005/mainAIED.pdf>.

[Carlson] Carlson, Bethany, and Sharon Reidy. "Effective Access: Teachers' Use of Digital Resources (Research in Progress)." <u>OCLC Systems & Services</u> 20 (2004): 65-70. <u>Emerald Library</u>. WPI Gordon Library, Worcester. 26 Aug. 2005 <http://www.emeraldinsight.com/1065-075X.htm>.

[Crisis] Capodilupo, Christina, and Anne Wheelock. "MCAS: Making the Massachusetts Dropout Crisis Worse." <u>MCAS Alert</u> 08 Sept. 2000: 3-6. <u>FairTest/CARE</u>. Education Resources Information Center. WPI Gordon Library. 26 Aug. 2005. Keyword: MCAS.

[Innovation] "Funding the Knowledge Economy: is the USA Losing Its Innovation Edge?" <u>Strategic Direction</u> 21 (2005): 33-35. <u>Emerald Library</u>. WPI Gordon Library, Worcester. 26 Aug. 2005 <http://www.emeraldinsight.com>.

[Koedinger] Koedinger, Kenneth R., Martha W. Alibali, and Mitchell J. Nathan. <u>A Developmental Model of Algebra Problem Solving: Trade-Offs Between Grounded and Abstract Representations</u>. Annual Meeting of the American Educational Research Association, 19 Apr. 1999, American Educational Research Association. 26 Aug. 2005 <http://www.eric.ed.gov>.

[Scaffolding1] Bull, Kay S., Paul Shuler, Robert Overton, Sarah Kimball, Cynthia Boykin, and John Griffin. <u>Processes for Developing Scaffolding in a Computer Mediated Learning Environment</u>. American Council on Rural Special Education, 25 Mar. 1999, Oklahoma State University. 26 Aug. 2005 <http://www.eric.ed.gov>.

[Scaffolding2] McLoughlin, Catherine. "Scaffolding: Applications to Learning in Technology Supported Environments." <u>ED-MEDIA 99 World Conference on Educational Multimedia and Hypermedia</u> (1999): 2-7. <u>Education Resources Information Center</u>. WPI Gordon Library. 26 Aug. 2005. Keyword: scaffolding.

[Transfer] Shariq, Syed Z. "How Does Knowledge Transform as It is Transferred? Speculations on the Possibility of a Cognitive Theory of Knowledgescapes." <u>Journal of Knowledge Management</u> 3 (1999): 243-251. <u>Emerald Library</u>. WPI Gordon Library. 26 Aug. 2005 <http://www.emeraldinsight.com>.

[Waks] Waks, Leonard, and Eric Sandelands. "The US Education System for the Non-American Faculty Member: an Interview with Leonard Waks." <u>Education + Training</u> 46 (2004): 33-39. <u>Emerald Library</u>. WPI Gordon Library, Worcester. 26 Aug. 2005 <http://www.emeraldinsight.com/10.1108/00400910410518205>

[Worcester] Beaman, Richard H., and Roberta R. Schaefer. <u>Benchmarking Public Education in Worcester</u>. Worcester Regional Research Bureau, MA. Center for Community Performance Measurement. Worcester, 2002. 26 Aug. 2005 <http://www.eric.ed.gov>.