

06D0401
NTH-CS37-51

Scaffolding in Computer Aided Learning

An Interactive Qualifying Project Report
submitted to the Faculty
of the
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the
Degree of Bachelor of Science
by
Derek Radtke
Hattie Spetla
Date: April 05, 2006

Approved:
Professor Neil Heffernan, Project Advisor

Abstract

The interactive online tutoring system Assistments was used to compare the effectiveness of online teaching using ‘scaffolding’ and a traditional hints only system for MCAS preparation. Students were tested in three local schools, Worcester North High School, Worcester Vocational High school, and Doherty High School. A control group was given hints, and the test group was given scaffolding. The student’s results were then analyzed to determine which was more effective.

Table of Contents

Abstract	2
Table of Contents	3
Introduction	4
Background	5
MCAS	5
The Assistments Project and Interactive Learning	6
Motivation for Scaffolding Questions vs. Hints	10
Motivation for Experimental Design	16
The Experiment	18
Purpose	18
Design	18
Content used	19
Comparison to past experiments	20
The Results	21
Analysis	22
Future Recommendations	38
Minor Changes	38
Take Home Assistments	39
Conclusion	41
References	42
Appendix A	43
Experiment 1	43
Experiment 2	44
Experiment 3	45
Appendix B	47
Appendix C	49
Appendix D	57
Acknowledgments	60

Introduction

In recent years many states have introduced standardized assessment tests to enforce required base levels of academic performance. Massachusetts introduced a set of such tests called the MCAS (Massachusetts Comprehensive Assessment System) tests. Many students struggle with these tests. To improve test scores, the Assistments project, a joint project between Worcester Polytechnic Institute and Carnegie Mellon University, provides automated tutoring to students in the Worcester area. We will study different ways of performing such tutoring in an attempt to improve how tutoring is performed.

There are two primary methodologies used in the Assistments system for teaching a student having difficulty with a question. The first is to provide hints, either generally about the problem such as subjecting which method to use to solve it or specifically pointing out why a chosen answer is wrong. The second is to expand incorrectly answered questions into a sequence of sub problems that ask about each specific facet of the main question, leading the student sequentially to the final, correct, answer. This use of sub problems is called “Scaffolding”. We would like to compare these methods to see which provides the greater benefit to the student.

To compare hints to scaffolding, we will perform a set of experiments using each method as a means of teaching and comparing the resultant change in student performance on tests before and after the tutoring. It is believed that the more in-depth scaffolding will lead to a greater degree of learning for the students. With this information, the Assistments system can be better used to provide tutoring to prepare students for standardized tests, especially the MCAS tests in Massachusetts.

Background

The MCAS (Massachusetts Comprehensive Assessment System) tests are used, in part, as a requirement for graduating high school in Massachusetts. The Assistment project is attempting to improve the passing rate of these tests through automated tutoring. There are several methods of providing computer tutoring, we will discuss the two types used by the Assistments system and provide justification for them. In addition, we will provide support for our method of comparing the different tutoring methods.

MCAS

Since 2003, MCAS tests have been administered annually to 3rd, 4th, 5th, 6th, 7th, 8th and 10th grade Massachusetts students. Scores are measured on a scale of 200 to 280 points. For the 10th grade tests, students are required to achieve a score of at least 220 points in both the English Language Arts section and the Mathematics section to graduate from high school. Therefore it is particularly important for 10th grade students to be familiar with and capable of completing the MCAS successfully. 10th grade students themselves are actively interested in learning the material presented because of this high stake.

These tests are used to rank the student's school academic performance. Therefore having all students pass the MCAS tests is also a point of great concern for the schools and their teachers. Worcester schools' 2005 32% failing rate in mathematics is worse than the state wide 15% mathematics failing rate further motivating the study of methods for improving student scores.

The MCAS math tests, which we base our study on, are designed to focus on four

areas of math competency: Number Sense and Operations; Patterns, Relations and Algebra; Measurement and Data Analysis; and Statistics and Probability. Number Sense and Operations focuses on basic arithmetic including simplification and the ability to use estimation to evaluate the results of problems using the real numbers. Patterns, Relations and Algebra includes understanding the varied representations of lines, manipulating polynomial and rational expressions, solving linear, quadratic equations with real roots, and inequalities, in addition to applying this to word problems taken from real life and finding patterns. Geometry primarily focuses on the properties of lines, angles, triangles, circles and 3 dimensional solids; including operations such as finding the slope or midpoint of a line. Finally, Data Analysis, Statistics and Probability includes problems ranging from interpreting data sets in graphical or tabular forms, approximating lines of best fit, and understanding the implications of sample size.

The Assistments Project and Interactive Learning

The Assistments project combines assisting with assessment. It is easy for schools to use the Assistments system as it is available as a web based application usable from any computer with a web browser and an internet connection. For a class to begin using the Assistments system, a school needs only to schedule the use of a computer lab for the class.

For the assisting, students are presented example problems drawn from previous MCAS tests to work on as they appeared on the MCAS tests. If the student becomes stuck or answers the problem incorrectly, the system either gives them a hint or presents a breakdown of the problem for them to work on. When a student is stuck, they can ask for a hint, which gives an overall suggestion of how to approach the problem. When presenting

an incorrect answer the system may also explain why that particular solution is incorrect. When there is not a simple explanation of the problem, the system presents a series of sub problems building up to the solution to the actual problem. These sub problems focus on the specific skills needed to solve the problem and lead the student through the problem.

For assessing, the system is capable of generating reports for teachers detailing each student or class' specific strengths and weaknesses. This enables teachers to focus their own teaching on preparing students for specific problem areas on the MCAS tests.

The MCAS tests four major areas of competency, as detailed in the section "MCAS"; these four sections are each too broad to focus teacher's efforts or find similarities between problems. To correct this failing, the developers of the Assistments system created lists of detailed 'Knowledge Components.' These knowledge components denote the more specific skills needed to solve the given question. Using these 'knowledge component' markings assigned to each actual MCAS items in the system, a teacher can tell which topics students are having the most trouble with.

Creating items for students to work with is easy for teachers and requires no programming skill. Questions, answers, hints and scaffolding can all be typed directly into a form like builder.

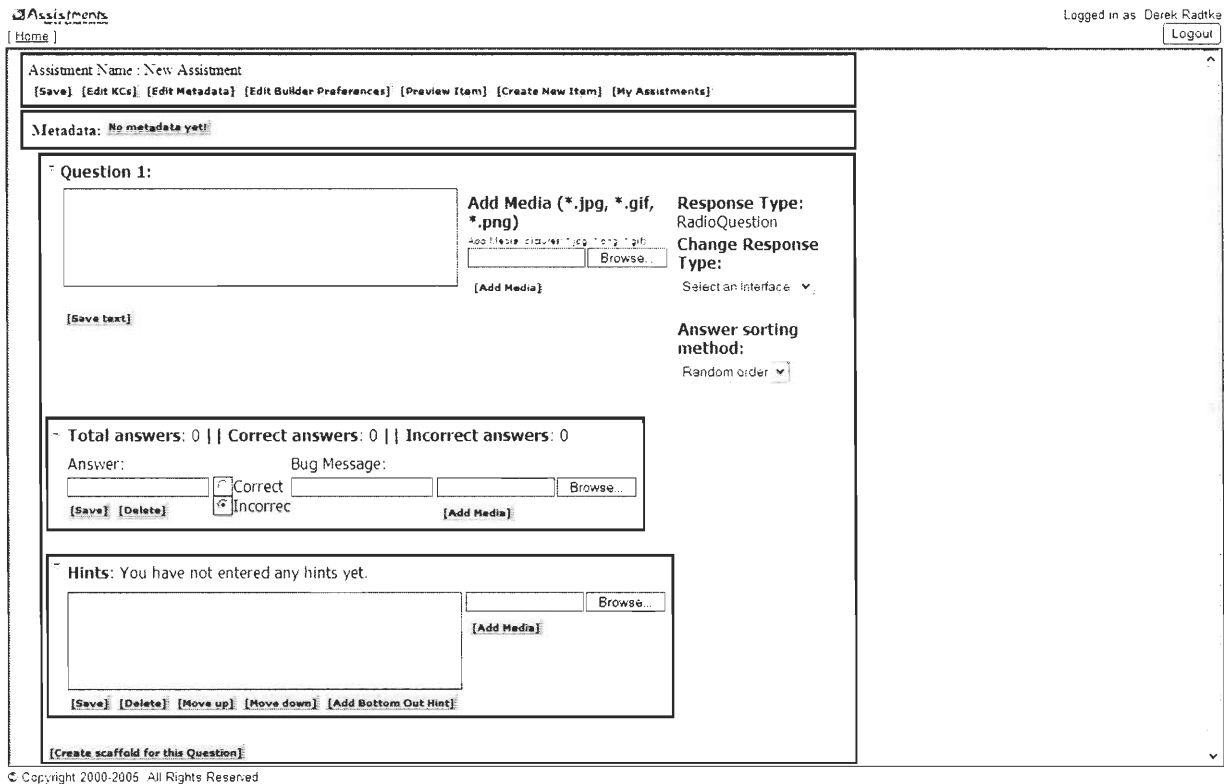
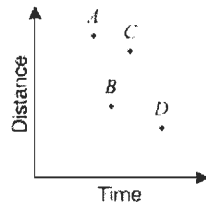


Figure 1.1: Assistsments Builder

For our project, before actually running an experiment, we created several dozen Assistsments items using the builder (see figure 1.1). The majority of the MCAS items we created were from the years 2001 and 2002. To help 10th grade students, we then created a set of scaffolding for each MCAS item. This scaffolding was a set of questions that focused on guiding the student through the problem. The number of scaffolding questions is typically equal to the number of knowledge components contained in the question. For example item 3 in the 2001 MCAS is shown in figure 1.2.

3 The scatter plot below gives information about four different car trips.



Which point represents the trip with the fastest average speed?

- ✓ A. point A
- B. point B
- C. point C
- D. point D

Figure 1.2: MCAS example

There are two skills necessary in completing this item, the ability to interpret points on a graph and also knowledge of how time and distance relate to speed. If a student incorrectly answered this problem, the system would ask them about which took the longest or went the furthest. At this level, explanations of how to read a scatter graph are available if the student requires them. Once the student had correctly read the scatter graph, the system would proceed to have the student determine which point had the largest distance per time and remind the student that distance per time is the speed.

Before beginning our experiment we created several dozen items for the system. Creating items gives a good impression of how the system works. We also gained a better understanding of benefits the Assistments system can have over other forms of online tutoring. Going to work with students and teachers at various Worcester public schools we determined what points of the system we would like to test. We also determined the most successful way for working with students. Taking the knowledge gained from working with the system we came up with our experimental design to test the effectiveness the items we created.

Motivation for Scaffolding Questions vs. Hints

How People Learn, published by the National Research Council, gives three specific aspects of successful learning environments. If the Assistments system can fit into each of these as set out by the National Research Council, then there is a very strong indication that learning will be seen in those students who use the system. When the system is strongly built to use important parts of the learning process, there is a basis for expecting student to learn better than a less developed system. The specific learning environments discussed are learner centered environments, knowledge centered environments and assessment centered environments.

A 'learner centered' environment takes into account what the student already knows. A diagnostic approach takes what the student already knows and finds misconceptions. The Assistments project attempts to find out whether or not a student has a certain skill. If not, it will go into greater detail to try and attempt to find out what part of the problem is giving the student difficulty. By not trying to teach the student things they already know, it is keeping in mind the personal skill of the learner, making it stronger program.

Another aspect of the learning environment is knowledge centered. A knowledge centered environment overlaps with a learner centered environment when the instruction starts with a consideration for the students preexisting knowledge. Learner centered environments also keep in mind what activities and teaching styles benefit students the most. Also, most importantly, these environments are centered on understanding as opposed to memorization. The Assistments system, in breaking problems down into smaller parts, helps to foster understanding in students. Understanding of the entire problem and the

ability to complete smaller tasks involved in the larger applies a knowledge based approach.

A successful learning environment also includes assessment and then feedback based on this assessment. Many studies cited by *How People Learn* have confirmed the importance of feedback in learning. With the goal of learning, assessment and feedback must be based on understanding the material presented. Simple grades, like tests and report cards are not so much instructive feedback, but more of a rating of the results of learning. Students do not receive grades until material is completed soon afterwards a new topic is started. In the Assistments system a student gets immediate and helpful feedback after every incorrect answer. They can also request help if they are completely unable to attempt to answer a question. This feedback fosters learning. The Assistments project assesses and gives feedback simultaneously, helping both the student and the teacher understand what is going on.

These are all strong reasons why the scaffolding system should help students learn more than a simple hints only system. The complexity of the scaffolding takes into account what the student already knows. If a student already knows how to answer a problem, no more work is done. If the student does not, the program attempts to figure out what part of the problem was giving the student trouble. It is knowledge centered. It does not try and get students to memorize, but to fully understand problems and the different smaller aspects that make up complex problem solving skills. Lastly, it is feedback centered. The scaffolding, unlike hints, gives students more complex feedback than a simple right or wrong. It trains them how to solve similar questions in the future and points out what things they are having difficulty with within the larger problem.

In Scaffolding vs. Hints in the Assistments System by Leena Razzaq, Neil T.

Heffernan, an early experiment testing the effectiveness of scaffolding in the Assistments system, performed in 2004 using 8th grade items. This experiment was very similar to our own and served as a starting point for many other experiments in the Assistments system. There were two conditions, one with scaffolding, one without. The paper describes a much higher gain in students who received the scaffolding treatment. There was a major problem, however, with this experiment. There was a higher rate of completed tests in the hints only group. The hints only group had much less material to go through than those given the scaffolding. There were fewer scaffolding tests completed, and those completed were more likely to have been done by better students, perhaps already familiar with the material, or harder working. These students could move much faster through the larger amount of scaffolding feedback than students less familiar with the material. This would, and most likely did, skew the results in favor of the scaffolding treatment.

To prevent this problem, Razzaq created a new experiment that was timed. For 20 minutes students worked on two pretest items and four experiment items. They were then given 15 minutes to complete another curriculum containing the 4 transfer items. This ensured that students in both conditions were completing the transfer items. There was a statistically significant increase in those post test questions that were seen in the pretest, with a p-value of .0005. Also, the scaffolding section did better than the hints only, unfortunately the result was not statistically significant.

In “When is Reading Just as Effective as One-on-One Interactive Human Tutoring?” by Kurt VanLehn, et al, three experiments were run to test the effectiveness of interactive human tutoring. If human tutoring is no more effective than interactive

computer based learning, the hope that interactive computer based learning is helpful to students is less likely. We can also learn from problems VanLehn et al had in the testing of the effectiveness of the tutoring and used this in our own work. They had mixed results.

In the first experiment there were two group. The first group was given reading based tutoring; the second group was given typing based tutoring. The group in the reading based tutoring read a detailed explanation of the correct answers to problems. Those in the typed condition engaged in typed tutoring with a human tutor. In this first experiment, there was no statistically significant difference in the scores of the two groups. The students tested had some familiarity with the subject matter. VanLehn hypothesized that because the students already had familiarity with the subject matter, they were only being refreshed in the material, which could be done equally well interacting with a tutor or reading a book.

The second experiment they performed was done with a group of students completely unfamiliar with the material being taught. They also formed a third test treatment. In addition to “Typed Interactive Tutoring”, students in a third group interacted one-on-one with a human tutor. For this experiment they found that but “Typed Tutoring” was more effective than reading (effect size 0.65) and “Spoken Tutoring” was much more effective than Reading (effect size 1.64). However this was for only one portion of the post test, a multiple choice test. An essay test showed no significant difference. Still, it's interesting to note the fact that more actual learning occurred with those unfamiliar with the material. Learning is what our experiment wants to determine. This is promising and can be used as a sign that interactive learning, like that of scaffolding in the Assistments system, is more useful in teaching students then a read only system. Reading only is very

similar to hints only problem in the system. A third experiment was also conducted. It involved shortening the experiment to lower the number of people who drop out. The results of the third experiment showed no significant difference between the test groups.

Several conclusions are drawn from this. Because texts cannot gauge the knowledge level of the student, it cannot lower the level of difficulty if the student is unfamiliar with the material. Interactive, in this case human, tutoring is much more effective in cases where students are completely unfamiliar with material. The Assistments system aims to be successful in this as well. Problems are broken down into smaller components.

In *e-Learning* by Ruth Clark and Richard E. Mayer, we are given guide lines for looking at similar trials. How similar are the learners in their study compared to our own? Are conclusions reached by an experimental research design? Are the experiment results replicated? Does the data analysis reflect statistical significance as well as practical significance? Looking at the previous two experiments, we can say that although the second one does not come as close to matching our design group as the first one, we can still draw valuable information from it. The students are around 4-5 years older than those in our study. However, they are still students and are similar in that respect to those in our study. Clark is referring to comparing student results to those of adults, many of whom are 'rusty' in learning.

In *Designing World-Class E-Learning*, Roger C. Schank's explanations of "Learning by Doing" support the hypothesis that the Assistments System's scaffolding should lead to higher learning than a hints only system. He argues that trial and error leads to the best results in learning; the Assistments system provides students the ability to

experiment with real MCAS problems and learn through direct experience. Further more, by breaking the problem down into its subcomponents, students can experiment with the sections of the problem they don't understand. He lists three typical problems with learning by doing. First, learning by doing can be dangerous. Making a mistake in the Assistments system in the "worse case" will be penalized by the teacher grading the student down. Secondly, learning by doing can be expensive; there are two types of possible cost, direct monetary costs and time costs. Since monetarily the schools are not directly paying for the service this is not an immediate concern. Time wise, Students are spending their normal class time with the Assistments System, so it is only of concern that the time spent with the Assistments System teaches better than that same time spent in normal class. Additionally the Assistments system can now be assigned as homework, further lower its cost. Finally, he warns that assuring the relevancy of example cases is important. The Asssitments problems are pulled from historical, real MCAS tests, the tests change very little year to year, so the problems remain relevant. This provides situations similar to what can and will occur. Hence the Assistments system is a good example of where to learn by doing as determined by Roger C. Schank's criteria.

Motivation for Experimental Design

This experiment focuses on how much students learned over the course of a curriculum. In *Evaluating E-Learning* William Horton calls an evaluation of the amount of learning a "Level 2 Evaluation". He raises a few points of concern on how testing the amount of learning may influence it. First, he warns that time limits may discourage students from exploring alternative paths which he believes is where a large component of learning lies; this experiment was always conducted in a single class period but it was designed to be short enough to be easily finished under such time constraints and the students were not made explicitly aware of the time limits. Secondly he is afraid that explicit tests also limit student's exploration. The system does allow such exploration giving the students multiple tries on any question it doesn't know scaffolding for. When it does have scaffolding, the scaffolding steps the students through the problem and allows them to experiment at each sub problem along the way. Either way the student gets to play with the problem until they get it correct during the learning portion of this experiment. Our experimental design also follows his recommendations of when to test. The experimental units each contain a pretest, his "beginning of the course" and the experiments are short and focused so our test at the end is equivalent to one of his intermediate tests embedded in the course. We did not look at the longer term learning like he would at the end of a course since our experiment was of a more limited scope. With a longer curriculum, an additional test after multiple sections would be required to follow Horton's suggestion of providing intermittent testing.

Roger C. Schank also points out that we learn from our failures, while Assistments do not design specific failure cases to occur as he suggests, it often is the case that problem

groups are assigned that focus on problem types students typically have difficulty with. Our curriculums in this experiment are designed to cover one such area each. He proposes that an advantage of e-Learning is that mistakes can be immediately explained by an expert. These experiment studies whether short concise instant explanations are better than detailed interactive explanations or visa versa. More detailed explanations provide a more complete description at the cost of being slower and hence extending the time between mistakes and teaching, a factor which Schank points out, is critical. Hints minimize this time delay but are less complete explanations and may miss the point of difficulty and do not allow experimentation.

The Experiment

Purpose

Our purpose in this experiment was determining the effectiveness of using scaffolding in the Assistments system's practice problems to improving the 10th grade MCAS scores of high school students. We hypothesized that the use of scaffolding Assistments would lead to greater learning as compared to hint message based Assistments. This hypothesis was based on the belief that more detailed exposure to the problems would aid comprehension and lead to greater learning. While some previous work had been done, this experiment was designed to avoid some possible flaws that had been found in the previous experiments and improve upon them.

Design

Our experiment was split into three sections and two conditions. The first group was condition 1 and received items with scaffolding. The second group was condition 2 and received items with hints only.

Both groups of students first receive the same pair of pretest items. The purpose of this was to gain a basis for assessing their knowledge of the material before they were tutored on it as a baseline. When students completed these pretest items, they then receive four experiment items depending on which condition they were in. These four items were the same questions, but depending on which condition group the students were in when they incorrectly answered the problem, they either received scaffolding walking them through the problem or simply received hints. Lastly, both conditions receive a four question post test containing identical items which we refer to as the 'transfer' items. Two

of these transfer items are identical to the pretest items. The other two post test items were new to the student, but are related to the topic.

Each item in the experiment was chosen to have very similar, if not identical knowledge components.

When performing testing, two tests were used in each class and were alternated by student such that no two students sitting next to each other had the same test. This was to limit communication between students on the test items so as to not artificially inflate some student's scores. In addition, all three sections followed immediately after the preceding section with no delay.

Content used

There were three separate experiment question sets developed. The first question set was developed to teach and test learning of triangle concepts, specifically focusing on the Pythagorean Theorem. This choice was made by the experimenters based on observation of what students seemed to have difficulty with. The later two experiments were on linear equations and percentages, chosen at the request of teachers.

The items used were obtained from those developed by this IQP group and the preceding IQP group consisting of Andrei Paduroiu and Emily Potter.

Comparison to past experiments

Previous experiments were conducted to determine the effectiveness of scaffolding versus hints in the Assistments system. However, these tests had several shortcomings our experiment tried to overcome:

- 1 Previous 10th grade experiments contained no pretest. This makes it especially difficult to track learning. With a pretest, those students who already know the material can be removed from the results before analysis.
- 2 The previous tests also had fewer conditioning questions. They did not focus on a specific topic. By focusing on specific topics, we hope to better gauge the effectiveness of the scaffolding. Using 4 condition questions instead of the only 2-3 condition questions on a previous experiment exposes the students to more of the treatment and should lead to a more measurable effect.
- 3 Using only 3 experiments, limited the number of variables influencing in the results. The previous experiment had 16 mini experiments.
- 4 This experiment takes into account time to compare effectiveness as relates to time.
- 5 Used multiple tests in an attempt to limit cheating in addition to monitoring which had been previously used.
- 6 The transfer followed immediately after the condition questions so as to not create an artificial gap between tutoring and testing for the hints students that was not as large as the gap for the scaffolding students. Previous experiments started the students on the transfer all at the same time, leading to students in the hints condition to have a greater delay in which to work on other unrelated problems than the students in the scaffolding condition due to the faster nature of hints.

The Results

Going to Worcester Vocational High, Worcester North high and Doherty we obtained several classes worth of information. Unfortunately, many of the results were unusable.

Some of the unusable data was due to classes being unable to finish in the allotted time. When time was too short for everyone to finish, those in the hints only group were more likely to finish than those in the scaffolding group, several of which would typically not complete the curriculum. This would leave incomplete data and bias that class's collected data through a selection effect. This would skew the data, so an entire class period of data had to be thrown out. This is doubly unfortunate, because that class could not be used again, as they had already seen the items in the experiment.

Some students got both pretest items right. These results are unusable, as well. Without incorrect pretest items, the post test results can only be equal to the pretest. There can be no indicator of learning.

Some students got all the experiment items right. By correctly answering all condition items, students were not exposed to neither hints or scaffolding. They were not exposed to the treatment, and therefore can not be considered to be in either experimental category.

Analysis

Initial Attempt

In initial examination of our results we observed something we referred to as negative learning. Negative learning consisted of many students who got the two pretest items correct subsequently getting the identical item incorrect in the post test. This was happening with an alarming and slightly scaffolding biased frequency.

We hypothesized two main reasons for this. The first reason may have been due to confusion at seeing the same question twice. While observing the experiment several students commented on the fact that the same question came up twice. The students have worked with the system before and know what kind of behavior to expect. Unexpected behavior from the system may cause the students to second guess a correct answer.

Another issue would have been the students' reaction to seeing the same question twice. While observing classes undergoing the experiment, many showed frustration at seeing the same question. One of the major problems with the Assistments system is student frustration. Seeing the same question twice, a student may be less likely to put in the effort to get the question correct the second time.

However, because these results were also out of sync with previous experiments, such as Razzaq's, we were motivated to double check the data. One of the more useful features of the Assistments system is an experimental 'Reporting tool.' This tool returns all results for a certain curriculum, regardless of school, teacher, or time taken. Without the tool we would have to go into the grade books of every different class who underwent the experiment and retrieve results manually. Results given by the tool are in a form which is easily modified to be

used in data analysis tools such as Statsview. Unfortunately the reporting tool was new and had never been used before. With this in mind, we went back to compare the raw results in a single grade book with those given to us from the 'Reporting tool.' As suspected there was a problem with the tool. It was not accurately reporting the correct results. Although these results were incorrect, we gained a little insight into the builder and spent time considering students reactions to the Assistments system. The problem raised serious concerns such as student frustration and confusion over items, and also pointed out a potentially dangerous problem with the reporting tool. We also learned a valuable lesson about checking and rechecking results.

Analysis of Results

We analyzed our results primarily using ANOVA tests in StatView a product of SAS Institute (<http://www.sas.com>). When comparing two similar samples, each sample often has a sizable overlap with the other. Since they are measured on the same scale, the min of one will likely be smaller than the max of the other, even if the samples do not contain these maxes and minimums the sample may simply be too small to contain them. Having only a limited sample, the max and min of each is likely to vary as would the mean, but the groups may or may not be distinct. Taking enough samples, you may conclude that if the mean differs enough, then the two groups are different. But how much is enough? A T-test allows us to compare two groups and see if it is likely they are the same or different. This works well when only two groups are compared, but as you add more groups, the number of comparisons grows rapidly. If you wish to conclude that there is an effect when

there is only a 5% or lower chance of a difference between two groups, and you compare 7 groups, there are 21 pairs and it would not be surprising for such a 1 in 20 chance to occur. The ANOVA test solves this problem by comparing all the groups together and generating a single probability of significance.

For this experiment we chose to use a p-value of 0.05 for acceptance. We generate our “gain score” from the number of problems correct in the transfer section that where in the pretest section minus the number of problems correct in the pretest section (all pretest items where repeated in the transfer section). Our transfer score is calculated as the number of items the student got correct in the transfer section. Participating in the condition is defined as the student having gotten at least one problem wrong in the condition section so that the student was exposed to either the hints or the scaffolding.

One Sample t-test
Hypothesized Mean = 0

	Mean	DF	t-Value	P-Value
gain score	.109	63	1.307	.1959

First we performed a t-test on our gain score to see if there was learning present in our overall data. While there was a mean increase of 0.109 problems correct on two problems, the p-value of 0.1959 indicates that it is quite possible this is merely a probabilistic occurrence and not a real effect. We then went and looked only at students who participated in the condition.

One Sample t-test
Hypothesized Mean = 0
Inclusion criteria: Exposed from all.txt (imported)

	Mean	DF	t-Value	P-Value
gain score	.117	59	1.308	.1961

This gave the slightly higher mean of 0.117 with a p-value of 0.1961, a slight improvement but still not near statistical significance. There was feedback given on the pretest items

which should have lead to some increase on its own, but the feedback was not clear, never explicitly stating whether the problem was correct or incorrect. The students have worked with the system before and know what kind of behavior to expect. It was believed that the unexpected behavior from the system may have caused the students to second guess their correct answer irregardless of whether it was correct or incorrect when they first gave it and possibly revert to guessing. Another issue that would also have been related to the student's reaction to seeing the same question twice is frustration. One of the major problems with the Assistments system is student frustration. Seeing the same question twice, a student may be less likely to put in the effort to get the question correct the second time. Many students asked why they where seeing the problem again a second time during the experiments, so we are not sure how this influence would affect the results.

Since overall there was not much learning, we looked at the learning in each experiment to see if there was significant learning in any of them.

One Sample t-test
Hypothesized Mean = 0
Inclusion criteria: Exposed to Exp1 from all.txt (imported)

	Mean	DF	t-Value	P-Value
gain score	.286	20	1.826	.0829

One Sample t-test
Hypothesized Mean = 0
Inclusion criteria: Exposed to Exp2 from all.txt (imported)

	Mean	DF	t-Value	P-Value
gain score	.063	31	.812	.4229

One Sample t-test
Hypothesized Mean = 0
Inclusion criteria: Exposed to Exp3 from all.txt (imported)

	Mean	DF	t-Value	P-Value
gain score	-.143	6	-.281	.7882

Experiment 1 had a mean increase of 0.286 problems correct on two problems with

a p-value of 0.0829, quite close to statistical significance. Experiment 2 had an increase of 0.0639 with a p-value of 0.4229, meaning it is likely that this increase was generated by a statistical fluke. The third experiment showed a decrease of 0.143 but it's p-value of 0.7882 was even worse than experiment 2's and lends little confidence to this being a real effect.

Given these results, the experiments were further compared to each other. First an ANOVA was performed comparing the pretest scores of students exposed to the experiment as determined by experiment.

ANOVA Table for pretest score

Inclusion criteria: Exposed from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Exp	2	9.727	4.863	9.510	.0003	19.019	.984
Residual	60	30.686	.511				

Means Table for pretest score

Effect: Exp

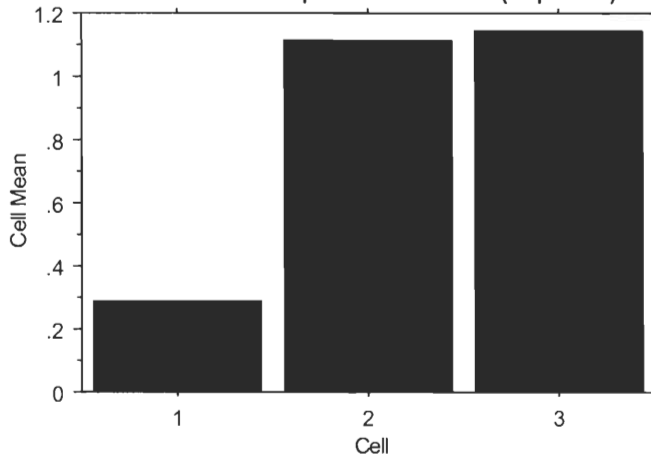
Inclusion criteria: Exposed from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	21	.286	.561	.122
2	35	1.114	.758	.128
3	7	1.143	.900	.340

Interaction Bar Plot for pretest score

Effect: Exp

Inclusion criteria: Exposed from all.txt (imported)



This showed experiment 1 having a mean pretest score of 0.286, experiment 2

having a mean pretest score of 1.114 and experiment 3 having a mean pretest score of 1.143. The tiny p-value of 0.0003 indicated it is very likely that these differences indicate an actual variance in pretests difficulty for the students. Experiment 1's pretest seems to have been by far the most difficult for the students. This gives students in experiment 1 more room to demonstrate improvement than students in experiments 2 or 3. This was followed by an ANOVA comparing the transfer scores between the experiments.

ANOVA Table for transfer score

Inclusion criteria: Exposed from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Exp	2	33.160	16.580	14.100	<.0001	28.200	.999
Residual	57	67.024	1.176				

Means Table for transfer score

Effect: Exp

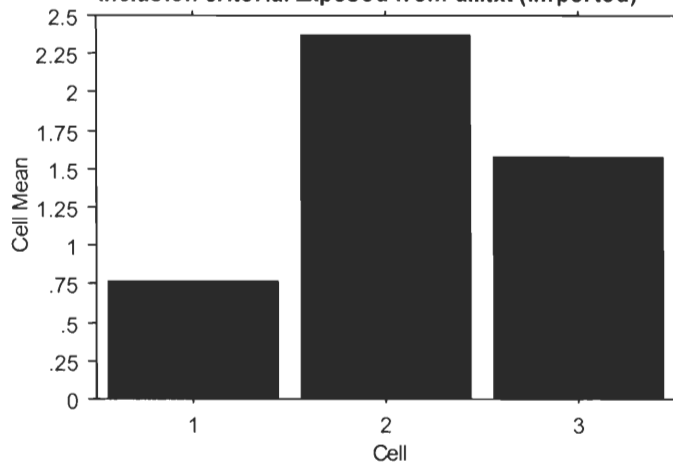
Inclusion criteria: Exposed from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	21	.762	.768	.168
2	32	2.375	1.185	.209
3	7	1.571	1.397	.528

Interaction Bar Plot for transfer score

Effect: Exp

Inclusion criteria: Exposed from all.txt (imported)



Experiment 1 was found to have a mean transfer score of 0.762, experiment 2 was found to have a mean transfer score of 2.375 and experiment 3 had a mean transfer score of

1.571. The p-value of less than 0.0001 leads us to the conclusion that this difference is almost guaranteed to be real. Experiment 1 persisted in being the hardest for students. To discount the possibility of this result coming from the difference in the pretest, an ANOVA was performed comparing the score on the two transfer items that students had not seen on the pretest by experiment.

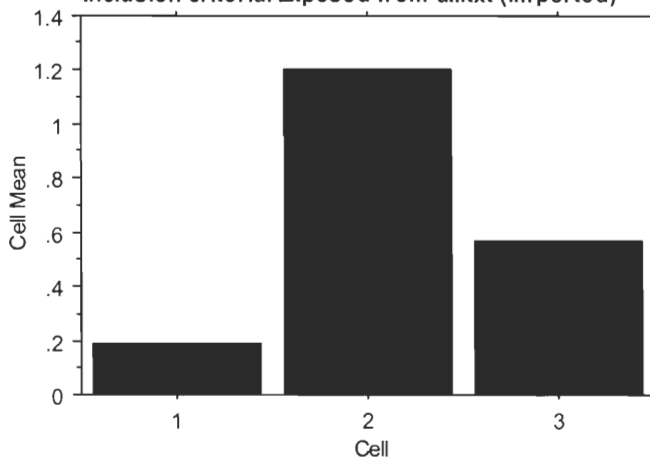
ANOVA Table for new transfer score
Inclusion criteria: Exposed from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Exp	2	13.765	6.883	16.819	<.0001	33.638	1.000
Residual	60	24.552	.409				

Means Table for new transfer score
Effect: Exp
Inclusion criteria: Exposed from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	21	.190	.402	.088
2	35	1.200	.719	.122
3	7	.571	.787	.297

Interaction Bar Plot for new transfer score
Effect: Exp
Inclusion criteria: Exposed from all.txt (imported)



Here it was found that experiment 1 had a mean score of 0.19, experiment 2 had a mean score of 1.20 and experiment 3 had a mean score of 0.571. This ANOVA also had a p-value of less than 0.0001, leading us to the conclusion these experiments really were

different and that they should be treated as such in further analysis.

Since we were interested in the difference between scaffolding and hints, an ANOVA test was performed with the gain score as the dependant variable with condition as the factor on all students who participated in the condition.

ANOVA Table for gain score

Inclusion criteria: Exposed from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	4.762E-4	4.762E-4	.001	.9751	.001	.050
Residual	58	28.183	.486				

Means Table for gain score

Effect: Condition

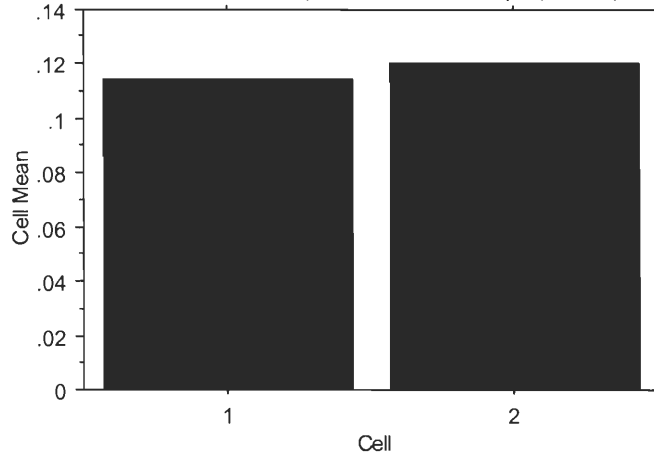
Inclusion criteria: Exposed from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	35	.114	.796	.135
2	25	.120	.526	.105

Interaction Bar Plot for gain score

Effect: Condition

Inclusion criteria: Exposed from all.txt (imported)



It was found that over all the experiment students in the scaffolding condition increased by 0.114 problems correct for two problems and that students in the hints condition increased by 0.12 problems for two problems but the p-value was 0.9751 which means these two results are unlikely to be different from each other.

While this might seem to indicate that differential learning had not occurred, the difficulty of the different experiments has been shown above to not be similar. Since we had seen that the experiments seemed to vary significantly from each other, we decided to also analyze each experiment separately.

ANOVA Table for gain score

Inclusion criteria: Exposed to Exp1 from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	.004	.004	.007	.9333	.007	.051
Residual	19	10.282	.541				

Means Table for gain score

Effect: Condition

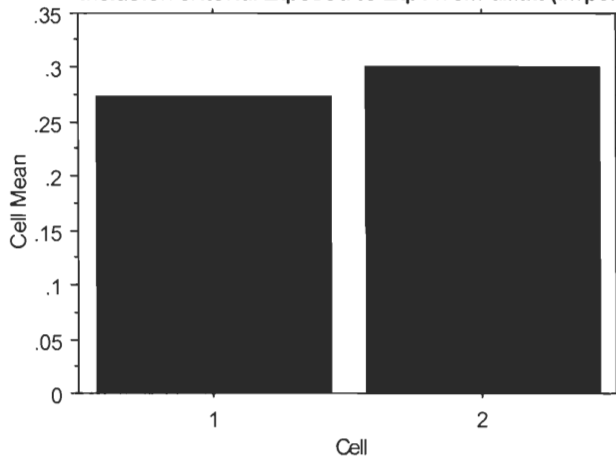
Inclusion criteria: Exposed to Exp1 from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	11	.273	.786	.237
2	10	.300	.675	.213

Interaction Bar Plot for gain score

Effect: Condition

Inclusion criteria: Exposed to Exp1 from all.txt (imported)



Experiment 1 did not come close to statistical significance with a p-value of 0.9333. The scaffolding condition showed a mean increase of 0.273 problems for two problems and the hints condition showed an increase of 0.3 problems for two problems but these results are not likely to actually be different given the p-value.

ANOVA Table for gain score

Inclusion criteria: Exposed to Exp2 from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	.110	.110	.574	.4546	.574	.110
Residual	30	5.765	.192				

Means Table for gain score

Effect: Condition

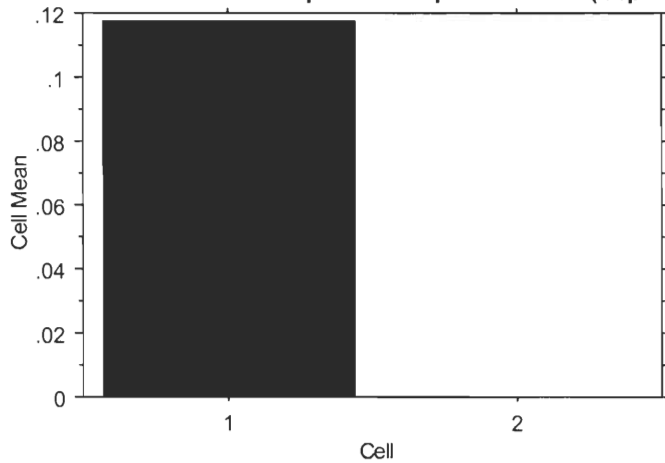
Inclusion criteria: Exposed to Exp2 from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	17	.118	.485	.118
2	15	0.000	.378	.098

Interaction Bar Plot for gain score

Effect: Condition

Inclusion criteria: Exposed to Exp2 from all.txt (imported)



Experiment 2 gave encouraging results for scaffolding, showing a mean increase of 0.118 problems for two problems for the scaffolding condition and no change for the hints condition. This did not reach statistical significance with a p-value of 0.4546. Experiment 3 lack sufficient students for an analysis and hence is omitted.

Having seen students confused by the repetition of problems from the pretest on the posttest, we considered that this might have influenced the gain score. There for we looked at the difference in the score on the remaining transfer items by condition.

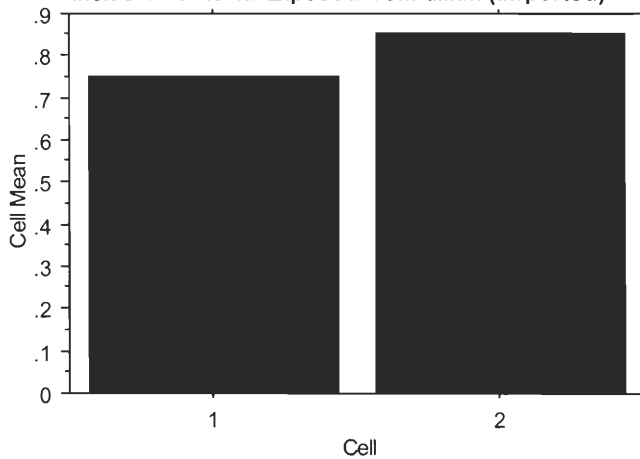
ANOVA Table for new transer score
Inclusion criteria: Exposed from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	.160	.160	.256	.6148	.256	.078
Residual	61	38.157	.626				

Means Table for new transer score
Effect: Condition
Inclusion criteria: Exposed from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	36	.750	.806	.134
2	27	.852	.770	.148

Interaction Bar Plot for new transer score
Effect: Condition
Inclusion criteria: Exposed from all.txt (imported)



Overall, the mean score of 0.75 for the scaffolding condition and of 0.852 for the hints condition are not significantly different with a p-value of 0.6148. Again thou, the experiments are not easily comparable due to the difference in difficult. Individually, some of the experiments fair a little better.

ANOVA Table for new transer score

Inclusion criteria: Exposed to Exp1 from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	.156	.156	.963	.3386	.963	.147
Residual	19	3.082	.162				

Means Table for new transer score

Effect: Condition

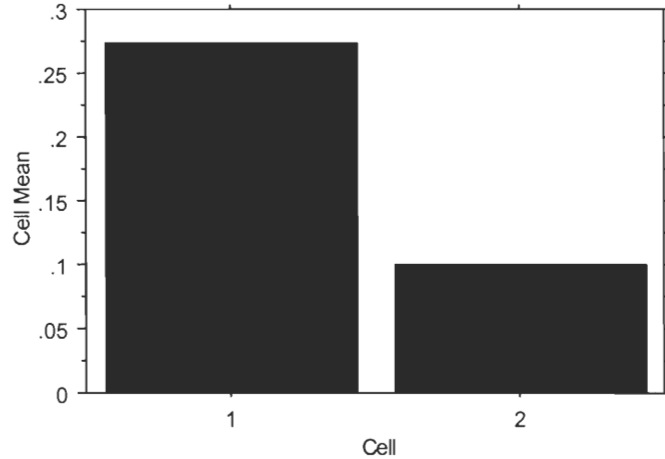
Inclusion criteria: Exposed to Exp1 from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	11	.273	.467	.141
2	10	.100	.316	.100

Interaction Bar Plot for new transer score

Effect: Condition

Inclusion criteria: Exposed to Exp1 from all.txt (imported)



Experiment 1's scaffolding condition had a mean score on these items of 0.273 and the hints condition had a mean score of 0.1. The p-value of 0.3386 is highly inconclusive thou.

ANOVA Table for new transer score
Inclusion criteria: Exposed to Exp2 from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	.293	.293	.558	.4602	.558	.109
Residual	33	17.307	.524				

Means Table for new transer score
Effect: Condition
Inclusion criteria: Exposed to Exp2 from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	18	1.111	.832	.196
2	17	1.294	.588	.143



Experiment 2's scaffolding condition had a mean score on these items of 1.111 while the hints condition had a mean score of 1.294 with the p-value of 0.4602, showing these results are also inconclusive.

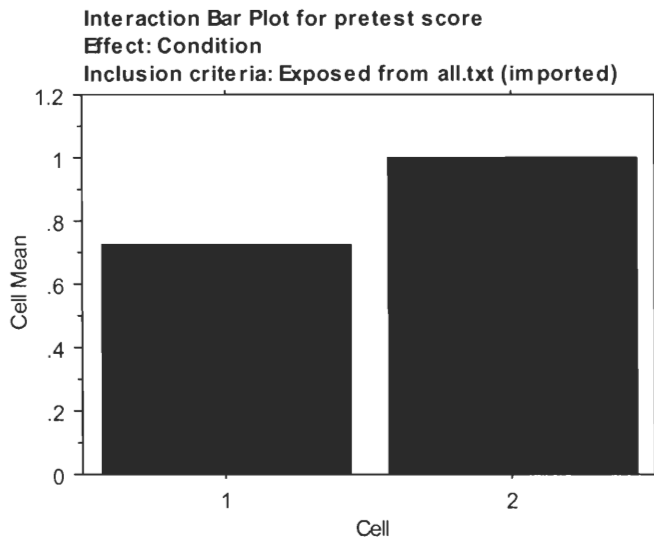
Thou most of these results where not statistically significant, a few of them seemed strange, so we tested to see if the conditions where balanced at the pretest.

ANOVA Table for pretest score
Inclusion criteria: Exposed from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	1.190	1.190	1.851	.1786	1.851	.253
Residual	61	39.222	.643				

Means Table for pretest score
Effect: Condition
Inclusion criteria: Exposed from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	36	.722	.815	.136
2	27	1.000	.784	.151



Overall, the scaffolding condition had a mean pretest score of 0.722 and the hints condition had a mean pretest score of 1. The p-value of 0.1786 is not statistically significant, but worrisome and this variance may still influence some of the above results.

ANOVA Table for pretest score

Inclusion criteria: Exposed to Exp1 from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	.004	.004	.012	.9147	.012	.051
Residual	19	6.282	.331				

Means Table for pretest score

Effect: Condition

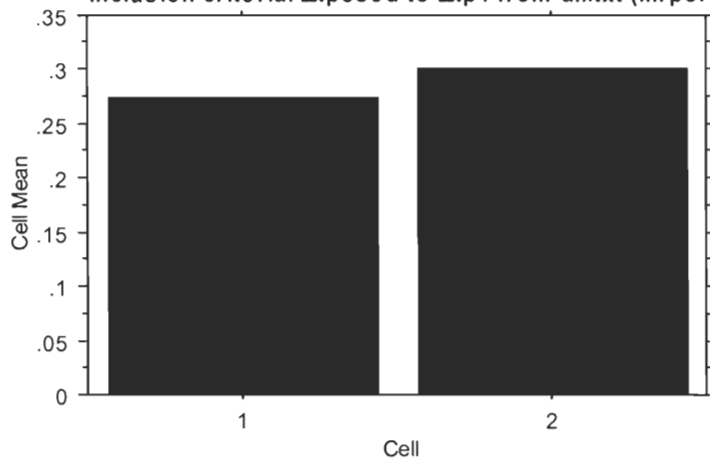
Inclusion criteria: Exposed to Exp1 from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	11	.273	.647	.195
2	10	.300	.483	.153

Interaction Bar Plot for pretest score

Effect: Condition

Inclusion criteria: Exposed to Exp1 from all.txt (imported)



Experiment 1 was much better balanced than the pretest scores overall. Experiment 1's scaffolding condition had a mean pretest score of 0.273 and its hints condition had a mean pretest score of 0.3 with a p-value 0.9147 making it fairly unlikely these are actually two different populations which is good for our analysis of the results based on this experiment.

ANOVA Table for pretest score

Inclusion criteria: Exposed to Exp2 from all.txt (imported)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Condition	1	2.925	2.925	5.809	.0217	5.809	.646
Residual	33	16.618	.504				

Means Table for pretest score

Effect: Condition

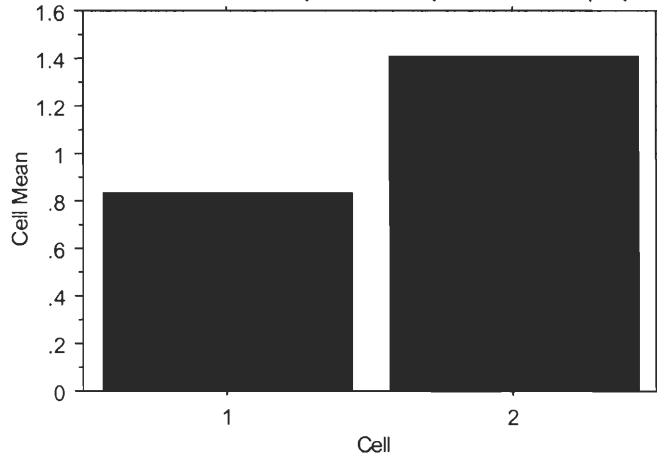
Inclusion criteria: Exposed to Exp2 from all.txt (imported)

	Count	Mean	Std. Dev.	Std. Err.
1	18	.833	.786	.185
2	17	1.412	.618	.150

Interaction Bar Plot for pretest score

Effect: Condition

Inclusion criteria: Exposed to Exp2 from all.txt (imported)



Experiment 2 on the other hand is highly unbalanced. Its scaffolding condition had a mean pretest score of 0.833 and it's hints condition had a mean pretest score of 1.412. The p-value of 0.0217 means that even thou condition was chosen at random these are likely to actually be different populations sampled and this strong bias very may well influence the results of experiment 2 including to improve the score of the hints condition on the transfer items they had not seen or increase their confidence on the pretest items repeated in the post test and hence limit the interference that the uncertain feedback may have generated.

Given the imbalance in the pretest of experiment 2 and the lack of data for experiment 3, the only usable experiment is experiment 1. This experiment was inconclusive given its p-values, but the new transfer score did lean towards the presence of differentiated learning between scaffolding and hints with scaffolding leading to a greater amount of learning. With further data, this may reach statistical significance.

Future Recommendations

Minor Changes

Some small changes to our experiment could make it more effective. More classes are being run. This will give more data, and most likely make more visible the difference between the learning rates of the two groups.

While data was collected on the time taken by students to complete the condition section of each experiment, the amount of data was too low to make this useful, further trials should make this data interesting. Given the large standard deviation in these times though, quite a few more trials are likely to be required before this data will be usable than simply comparing the amount learnt.

Longer condition sections would increase the effect, a mere four problems provides little opportunity for students to hone their skills. While they may be corrected once on a mistake for example, they are unlikely to have a chance to apply this new knowledge and test it. The difficulty with this is that logistics and availability of students limits the time the experiment can be conducted in and extending the experiment to cover multiple days is confounded by external teaching in class.

The assumption is made that short term knowledge gain is an indicator of long term

learning. While this seems a straight forward assumption, it could lead to slight differences between the results and reality even when correct. To remove this assumption the transfer section could be administered the next day. To avoid external teaching, no homework should be assigned on that night and no math classes should occur between the sections.

The pretest items could be made harder in future experiments. This would limit unusable test results stemming from students who appear as initially knowing the material.

Another issue with the reporter was that only problem names are reported, not problem numbers. Some problems were used multiple times in this experiment, such as problems in the pretest and posttest. While each of these problems had different problem IDs, the problem names were occasionally identical, hence processing the reports into a format that Statview could manage required de-obfuscating the problem names. This should be fixed in future revisions.

Take Home Assistments

Another possible experiment for future groups is to test the effectiveness of at home use of the Assistments system. There are many conditions that could be tested for in the at home environment that can't be determined in the classroom environment. There would also be other advantages to testing at home.

First, students would have less opportunity to cheat via communication with their peers, a method found to be common in the class room. This has been a problem in previous experiments and was found difficult to prevent even with multiple test curriculums. Even with modern communication cheating requires more explicit action than noticing someone else's screen. Since the difficulty of initiation is higher it should be more

infrequent. Further more, unlike a classroom setting, students would not all be working on the assignment at the same time or available to each other.

Secondly more students could be reached. Administration of this experiment caused difficulty since classes had to be attended. With a take home experiment, all students in a school could be reached without requiring sufficient experimenters to attend each class.

Conclusion

The Assistments system has a very promising future. Although we were unable to provide any directly useable results, the willingness of student and teachers to work with the system, the relative ease of which experiments and items could be created and run are all good indicators that the system will be determined to be effective in further experiments.

As more data for our experiments 1 and 2 is collected, these results should reach statistical significance. Experiment 1 is the most promising, but one could compensate for the imbalance in the pretest of experiment 2, or the imbalance should dissipate with increased data. With so few data points, the statistics are still easily affected by small chances in the data. Over all the results are promising, but inconclusive.

References

- [1] *Mathematics Curriculum Framework* (pages 72-75).
www.doe.mass.edu/frameworks/math/2000/final.pdf
- [2] *MCAS Website*, <http://www.doe.mass.edu/mcas/>
- [3] Andrei Paduroiu, Emily Potter, *Effective Teaching Methods: Are Scaffold Questions Better Than Hints?*
- [4] ANOVA: ANalysis Of VAriance between groups, <http://www.physics.csbsju.edu/stats/anova.html>
- [5] Feng, M., Heffernan, N. T., Koedinger, K. (2005) *Predicting State Test Scores Better with Intelligent Tutoring Systems: Developing Metrics to Measure Assistance Required*. Submitted to the 8th International Conference on Intelligent Tutoring Systems.
http://nth.wpi.edu/pubs_and_grants/ITS2006/Submissons/Ming/feng.doc
- [6] Razzaq, L., & Heffernan, N. T. (In submission). *Scaffolding vs. Hints in the ASSISTment System*. Submitted to the 8th International Conference on Intelligent Tutoring Systems (2006).
http://nth.wpi.edu/pubs_and_grants/ITS2006/Submissons/Leena/razzaq.doc
- [7] VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2005). *When is reading just as effective as one-on-one interactive human tutoring?* *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 2259-2264). Mahwah, NJ: Erlbaum.
<http://www.psych.unito.it/csc/cogsci05/frame/poster/3/f801-vanlehn.pdf>
- [8] Clark, Ruth. Mayer, Richard 2003 John Wiley and Sons San Fransisco. *e-Learning and the science of Instruction*.
- [9] Rodger C. Schank, Ph.D., 2002 McGraw-Hill. *Designing World-Class E-Learning*
- [10] William Horton, 2003 The American Society for Training & Development. *Evaluating E-Learning*
- [11] National Research Council, 2000 National Academies Press. *How People Learn: Brain, Mind, Experience, and School*

Appendix A

Experiment 1

Curriculum Name Six item learning section on slopes

Curriculum ID 1206

Pretest

pretest_triangle_2004_33: *A ladder to a roof is quite the hypotenuse*

Using the measures shown in the sketch, what is the length of the section of the ladder from the point where it rests on the ground to the point where it touches the house?

pretest_triangle_2002r_40: *Tall support for a thin line*

In order to stabilize a tightrope at a circus, a wire that has a length of 30 feet is attached from the top of the vertical support at point A to point B on the ground. Point B is 10 feet from the base of the vertical support as shown in the figure above.

Based on this information, what is the value of h to the nearest whole number?

Experiment Section

exp1_2001_32: *x is the height of 1 triangle and a side in another*

What is the length x of the altitude of the triangle shown above?

exp1_2001item17: *TriangleKite*

It is believed that the best angle to fly a kite is 45 degrees. If you fly a kite at this angle and let out 225 feet of string, approximately how high above the ground will the kite be?

E1_2004r25:

The lengths of the legs of a right triangle are 5 centimeters and 10 centimeters. Which of the following measures is closest to the length of the hypotenuse?

E1_2000t37:

A 25-foot wire attached to an antenna makes a 30° angle with the level ground, as shown above. What is the approximate distance from the base of the antenna to the place where the wire is staked to the ground? (The figure is not drawn to scale.)

Transfer/Post Test Section

2002r-10-26:

A boat traveled 4 miles due east away from a dock. Then it turned and traveled 5 miles due north. Finally, it turned again and traveled back to the dock as shown in the figure above. Dock 4 miles Which of the following is closest to the total distance the boat traveled?

T1_2004n40:

Triangle ABC, shown below, is a right triangle. What is the length of AB?

2002r-10-40:

In order to stabilize a tightrope at a circus, a wire that has a length of 30 feet is attached from the top of the vertical support at point A to point B on the ground. Point B is 10 feet from the base of the vertical support as shown in the figure above. Based on this information, to the nearest whole number, what is the value of the height of the vertical support?

exp_2004_33:

Using the measures shown in the sketch, what is the length of the section of the ladder from the point where it rests on the ground to the point where it touches the house?

Experiment 2

Curriculum Name Six item learning section on slopes
Curriculum ID 1226

Pretest**2004re37_exp:**

The students at a new high school voted for school colors. The bar graph represents the percent of students who voted for each color combination. Based on the graph, if 72 students voted for Purple and Gold, how many students voted for Red and Black?

2002r-10-11: 2002r-10-11

Oldwick High School has a population of 2345 students. A survey showed that 34% of the student population would like to have hamburgers as a daily choice for lunch. Which of the following is closest to 34% of the student population?

Experiment Section**1999_14: Percentage of O blood**

A survey showed that the distribution of blood types among people with a positive Rh factor is the following:

Type O.... 46%

Type A.... 39%

Type B 11%

Type AB.... 4%

If there are 750 students with a positive Rh factor in Martin High School, how many of these students would you expect to have Type O blood?

2001item18: 10th grade percentages

This table shows the distribution of CD sales of different types of music in a local store over a three-year period. Based on this information, from Year 1 to Year 3 the total dollar value of this store's sales of pop music

E2_2005t16:

On an airline, approximately 10% of the airline passengers who are booked for a flight do not show up for the flight. The airline has booked 160 passengers for a flight with maximum seating of 135. How many of the 160 passengers booked for this flight will not have a seat, assuming 10% of the booked passengers do not show up?

2002-g10-25:

The circle graph above shows the Corbett family's monthly budget. The Corbett family has a total monthly income of \$2,000. Mr. Corbett received a \$100 per month raise. He increased the transportation expense portion of the monthly budget by \$100. To the nearest percent, what portion of the Corbetts' income is now being spent on transportation expenses?

Transfer/Post Test Section

2005_37_gr10: *leena/Assistments/2005_37_gr10/2005_37_gr10-problem0.xml*

The chart above separates the number of students majoring in math/science from students pursuing other majors at a state college.

What percent of the math/science majors are seniors?

2005rMarch3:

Tomika collected data about the types of transportation that the juniors and seniors at her high school use to get to school. This table shows the data that she collected. Based on the data in the table, which of the following is closest to the percent of juniors and seniors combined who take the bus to school?

exp2_trans_2004R_37:

The students at a new high school voted for school colors. The bar graph shown above represents the percent of students who voted for each color combination. Based on the graph, if 72 students voted for Purple and Gold, how many students voted for Red and Black?

2002r-10-11: *2002r-10-11*

Oldwick High School has a population of 2345 students. A survey showed that 34% of the student population would like to have hamburgers as a daily choice for lunch. Which of the following is closest to 34% of the student population?

Experiment 3

Curriculum Name Six item learning section on slopes

Curriculum ID 1266

Pretest

linpretest-2005-10:

Which of the following best represents the equation of the line shown on the graph above?

linpretest-2005-35:

The scatterplot above shows the ages and heights of 20 trees on a tree farm. If x = age in years and y = height in meters, which of the following equations best approximates the line of best fit for this scatterplot?

Experiment Section

2002r-10-09:

Which of the following best represents the equation of the line shown on the graph below?

2002-g10-10:

Use the graph below to answer the question. Which of the following equations best represents the data in the graph?

2003_LineBestFit: *Finding the equation of the line of best fit in a scatterplot*

Jenny studied the effect of light on plant growth. She graphed a scatterplot to represent her data.

Which of the following **best** represents the equation for the line of best fit for the data shown?

2003_Retest_10g_18:

Write an equation for the line drawn above.

Transfer/Post test Section

2005-g10-10:

Which of the following best represents the equation of the line shown on the graph above?

2005-g10-35:

The scatterplot below shows the ages and heights of 20 trees on a tree farm. If x = age in years and y = height in meters, which of the following equations best approximates the line of best fit for this scatterplot?

2003r-g10-26:

The scatterplot above shows the length and mass of 10 pencils.

If x = length in centimeters and y = mass in grams, which of the following equations most closely approximates the line of best fit?

2004r-g10-7:

Which of the following is the apparent slope of the line graphed below?

Appendix B

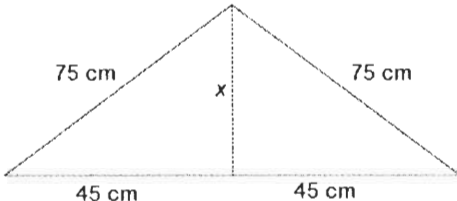
Data after being run through processor:

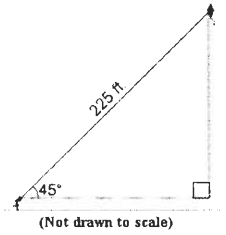
"UID", "p1", "p2", "te0", "e1", "e2", "e3", "t0", "t1", "t2", "t3", "TimeTaken", "Condition", "Exp"

"26587", "0", "0", "0", "0", "1", "0", "0", "0", "0", "0", "264", "2", "1"
"26590", "0", "1", "1", "1", "1", "0", "0", "0", "0", "0", "445", "2", "1"
"26102", "0", "0", "1", "1", "0", "0", "0", "1", "0", "0", "618", "1", "1"
"21597", "1", "1", "1", "0", "1", "0", "0", "1", "0", "1", "1270", "1", "1"
"26578", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "264", "2", "1"
"26596", "0", "0", "0", "0", "0", "0", "0", "1", "0", "0", "888", "1", "1"
"21933", "0", "1", "1", "0", "1", "0", "1", "1", "0", "1", "442", "2", "1"
"26580", "1", "0", "0", "1", "0", "0", "0", "0", "0", "1", "954", "1", "1"
"26635", "0", "0", "1", "1", "1", "0", "1", "0", "0", "0", "708", "2", "1"
"26599", "0", "0", "0", "0", "0", "1", "0", "0", "0", "0", "1164", "1", "1"
"26588", "0", "0", "0", "0", "1", "0", "0", "1", "0", "0", "339", "2", "1"
"26630", "0", "0", "1", "1", "1", "0", "0", "1", "0", "0", "927", "1", "1"
"26581", "0", "0", "0", "0", "1", "1", "0", "0", "0", "0", "942", "1", "1"
"26591", "0", "0", "1", "0", "1", "1", "0", "0", "0", "0", "812", "1", "1"
"26592", "0", "1", "0", "1", "0", "0", "0", "1", "0", "0", "274", "2", "1"
"26111", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "1512", "1", "1"
"26595", "0", "0", "0", "1", "0", "1", "0", "1", "0", "0", "1320", "1", "1"
"21935", "0", "0", "1", "1", "0", "1", "1", "0", "0", "0", "657", "1", "1"
"26584", "0", "0", "0", "0", "1", "0", "0", "1", "0", "0", "913", "2", "1"
"21598", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "147", "1", "1"
"26109", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "612", "2", "1"
"26597", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "674", "2", "1"
"26107", "0", "0", "0", "0", "0", "0", "0", "0", "0", "1", "910", "1", "2"
"26587", "0", "1", "0", "0", "0", "0", "0", "1", "0", "1", "405", "2", "2"
"26110", "1", "1", "1", "0", "1", "1", "1", "1", "0", "1", "576", "1", "2"
"26586", "1", "0", "0", "0", "0", "0", "0", "0", "0", "1", "1", "352", "2", "2"
"26101", "0", "0", "0", "0", "1", "0", "0", "1", "1", "1", "813", "1", "2"
"25637", "1", "1", "1", "0", "1", "1", "1", "1", "1", "1", "359", "2", "2"
"21605", "0", "1", "0", "0", "0", "0", "0", "1", "1", "1", "631", "2", "2"
"26593", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "1569", "1", "2"
"25638", "0", "0", "1", "1", "1", "1", "0", "0", "1", "0", "339", "1", "2"
"25635", "0", "1", "1", "1", "1", "1", "0", "1", "1", "0", "659", "1", "2"
"26598", "0", "1", "0", "0", "0", "0", "0", "1", "0", "0", "1607", "1", "2"
"26583", "1", "0", "1", "0", "0", "0", "0", "1", "0", "1", "1", "1059", "1", "2"
"26626", "1", "0", "0", "0", "0", "0", "0", "0", "0", "0", "1336", "1", "2"
"21939", "1", "1", "1", "0", "1", "0", "1", "1", "1", "0", "277", "2", "2"
"25640", "0", "1", "1", "0", "1", "0", "0", "1", "1", "1", "315", "2", "2"
"26881", "1", "1", "1", "0", "1", "0", "1", "1", "1", "1", "668", "2", "2"
"26589", "0", "0", "0", "1", "0", "0", "0", "0", "0", "0", "704", "1", "2"
"26578", "0", "0", "1", "1", "0", "1", "0", "0", "0", "0", "397", "1", "2"
"26596", "0", "0", "0", "0", "0", "1", "0", "0", "1", "0", "379", "2", "2"
"25632", "1", "1", "0", "1", "1", "1", "0", "1", "1", "0", "352", "2", "2"
"25639", "0", "1", "1", "0", "0", "0", "1", "1", "1", "0", "458", "2", "2"
"25641", "1", "1", "0", "0", "0", "1", "1", "1", "1", "0", "356", "2", "2"
"26456", "0", "1", "1", "0", "0", "0", "0", "1", "1", "1", "611", "1", "2"
"26585", "1", "1", "1", "1", "0", "0", "1", "1", "1", "0", "799", "2", "2"
"25645", "0", "1", "1", "0", "1", "0", "0", "1", "1", "0", "519", "2", "2"
"21940", "1", "0", "0", "0", "0", "0", "1", "0", "1", "0", "580", "1", "2"
"26579", "1", "1", "1", "0", "1", "0", "1", "1", "0", "0", "263", "2", "2"
"26108", "0", "0", "0", "1", "1", "0", "0", "0", "1", "1", "1421", "1", "2"
"26592", "1", "1", "1", "1", "0", "1", "1", "1", "1", "1", "394", "1", "2"

"21943", "1", "0", "0", "1", "0", "1", "1", "0", "1", "0", "483", "2", "2"
"25636", "1", "1", "0", "1", "1", "0", "0", "1", "1", "1", "956", "1", "2"
"25642", "0", "1", "1", "0", "1", "0", "1", "1", "1", "0", "596", "1", "2"
"21598", "1", "0", "1", "0", "0", "0", "1", "0", "1", "0", "650", "2", "2"
"26584", "0", "1", "0", "1", "1", "0", "1", "1", "1", "1", "521", "1", "2"
"26757", "1", "1", "1", "0", "0", "0", "1", "1", "1", "1", "325", "2", "2"
"25644", "0", "0", "0", "0", "0", "1", "0", "0", "0", "1", "1037", "1", "2"
"21607", "1", "1", "1", "0", "0", "0", "1", "1", "1", "0", "754", "1", "2"
"25940", "0", "1", "0", "0", "0", "0", "0", "1", "0", "0", "1213", "1", "3"
"25646", "1", "1", "1", "0", "0", "1", "1", "1", "0", "1", "386", "1", "3"
"25932", "1", "0", "0", "1", "0", "0", "0", "0", "1", "0", "2065", "1", "3"
"25936", "1", "1", "0", "0", "0", "0", "1", "0", "0", "0", "779", "1", "3"
"25640", "1", "1", "1", "1", "0", "0", "0", "0", "0", "0", "922", "1", "3"
"25939", "1", "1", "1", "1", "1", "1", "1", "1", "0", "1", "724", "1", "3"
"26757", "0", "0", "0", "0", "1", "0", "1", "0", "0", "0", "789", "1", "3"
"25632", "0", "0", "1", "1", "0", "0", "1", "1", "1", "1", "708", "1", "3"

Appendix C

Item Name	Problem Text
2001, p32	<p>Problem text:</p>  <p>What is the length x of the altitude of the triangle shown above?</p> <p>Breakdown using scaffold questions:</p> <ol style="list-style-type: none"> Thinking of x as the side of a new triangle, like the one above, makes this problem easier to solve. This new triangle is what type of triangle? <ul style="list-style-type: none"> The altitude, x, forms a 90 degree angle with the base. A 90 degree angle is also known as a right angle. That makes this a right triangle. Choose C. Now that you know you have a right angled triangle, you can use the Pythagorean Theorem to solve for x. Look for the equation in the reference sheet whose link is given in the upper right hand corner of the page. In the Pythagorean Theorem, the variable c represents the hypotenuse of the triangle. What is the value of c in the triangle above? <ul style="list-style-type: none"> The hypotenuse is the side opposite the right angle. In this case 75 is the hypotenuse. Type 75. Good. You know that $75^2 = a^2 + b^2$. a and b represent the two non-hypotenuse sides of the triangle. Plug 45 in for a and solve for b. Do you now know the length x of the triangle above? <ul style="list-style-type: none"> $75^2 = 45^2 + b^2$, where b represents x. Solve for b. $75^2 = 75 * 75 = 5625$ and $45^2 = 45 * 45 = 2025$ so $5625 = 2025 + b^2$. $3600 = b^2$. Square both sides. The square root of 3600 is 60. Type 60. <p>Breakdown using hints:</p> <ol style="list-style-type: none"> Try thinking of x as the side of a new triangle like the one above. The altitude forms a 90 degree angle with the base. Now that you know you have a right angle triangle, you can use the Pythagorean Theorem to solve for x. In the Pythagorean Theorem, what does the variable c represent? c represents the hypotenuse, the side opposite the right angle. a is one of the sides, so you can plug 45 into the equation for a. You now know that $75^2 = 45^2 + b^2$. $b = x$ $75^2 = 75 * 75 = 5625$ $45^2 = 45 * 45 = 2025$ $5625 - 2025 = b^2$ $3600 = b^2$ Take the square root of both sides. The square root of 3600 is 60. x is equal to 60.

Problem text:

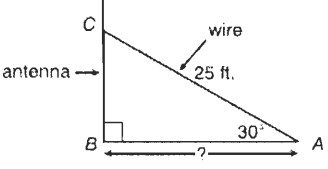
It is believed that the best angle to fly a kite is 45 degrees. If you fly a kite at this angle and let out 225 feet of string, approximately how high above the ground will the kite be?

Breakdown using scaffold questions:

1. What is the value the top angle of the triangle?
 - All the angles in a triangle must add up to 180.
 - The square drawn in the bottom right corner means that the angle is 90 degrees.
 - $90+45=135$ and $180-135=45$. The angle must be 45 so that all the angles add up to 180. Choose C.
2. Because this is a 45-45-90 triangle, the two sides opposite the 45 degree angles must be equal. You can use the Pythagorean theorem to solve for the height of the kite. The Pythagorean Theorem states $a^2 + b^2 = c^2$. What is the value of c for this triangle?
 - 1 C represents the hypotenuse, or the angle opposite the right angle. In this case it's 225, so type 225.
3. Good, now knowing that $c=224$ and $a=b$, approximately how high above the ground will the kite be?
 - 2 $225^2 = a^2 + b^2$. But because $b=a$, you can also write this equation as $225^2 = a^2 + a^2$.
 - 3 $225^2 = 2a^2$. $225^2 = 225 * 225 = 50625$
 $50625 = (2a^2)$. Divide 2 from each side. $25312.5 = a^2$
 - 4 Take the square root of each side. The square root of 25312.5 is equal to 159.1. This makes each side of the triangle, including the height, 159.9. This is closest to C. Choose C.

Breakdown using hints:

1. The angles of a triangle add upto 180 degrees.
2. The square in the bottom right angle means that that angle is a 90 degree angle.
3. Because this is a 45-45-90 triangle, both sides opposite the 45 degree angles must be of the same length.
4. The Pythagorean Theorem states that $a^2 + b^2 = c^2$.
5. Since $a=b$, you can write the Pythagorean Theorem as $a^2 + a^2 = c^2$.
6. The square root of 25312.5 is 159.1
7. The answer is 150 feet. Please enter 150 feet.

<p>2000, p37</p>	<p>Problem text:</p>  <p>A 25-foot wire attached to an antenna makes a 30° angle with the level ground, as shown above. What is the approximate distance from the base of the antenna to the place where the wire is staked to the ground? (The figure is not drawn to scale.)</p> <p>Breakdown using scaffold questions:</p> <ol style="list-style-type: none"> We have the hypotenuse, denoted as c, of a 30-60-90 triangle, and are looking for the length of the long leg, denoted as a. Select the formula to use from below. <ul style="list-style-type: none"> Check your formula sheet to find the relations between the sides. Now that you have a formula, what is a? <p>Breakdown using hints:</p> <p>In response to answers:</p> <ul style="list-style-type: none"> This is the height of the antenna, try solving for the other leg of the triangle. The wire its self is only 25 feet, if you laid it along the ground, it could not reach more then 25 feet. <ol style="list-style-type: none"> It is a 30-60-90 triangle, try looking at your formula sheet.
<p>2004, p25</p>	<p>Problem text:</p> <p>The lengths of the legs of a right triangle are 5 centimeters and 10 centimeters. Which of the following measures is closest to the length of the hypotenuse?</p> <p>Breakdown using scaffold questions:</p> <ol style="list-style-type: none"> For this problem, you want to use the Pythagorean Theorem. Which of the following equations is the Pythagorean Theorem? (you can check your reference sheet!) The legs of the triangle represent which two variables in the Pythagorean Theorem? After plugging in the values for the legs and solve the right hand side of the equation, you get: Now that we have solved most of the equation, we are ready for the last step, approximately how long is c (the variable representing the length of the hypotenuse)? <p>Breakdown using hints:</p> <ol style="list-style-type: none"> Use the Pythagorean Theorem, check your equations sheet. The legs are represented as a and b.

<p>1999, p14</p>	<p>Problem text: A survey showed that the distribution of blood types among people with a positive Rh factor is the following: Type O.... 46% Type A.... 39% Type B 11% Type AB.... 4% If there are 750 students with a positive Rh factor in Martin High School, how many of these students would you expect to have Type O blood?</p> <p>Breakdown using scaffold questions:</p> <ol style="list-style-type: none"> Percent means 'per hundred.' It's a way to express a ratio using whole numbers. In this example, 46 per 100 (46/100) students have type O blood. We need to know how many students, which we will represent with the variable x, out of 750 have type O blood. To find out how many out of 750 have type O blood, we can set 46/100 equal to the ratio $x/750$. First, what is 46/100? Good, now set 0.46 equal to $(x/750)$. Remember, x is the number of students out of 750 that have type O blood. $.46 = x/750$ The value of x should tell you how many of the students would you expect to have Type O blood. What is the number of students that would have type O blood? <ul style="list-style-type: none"> Multiply each side of the equation by 750. $.46 * 750 = 345$ Enter 345 <p>Breakdown using hints:</p>																												
<p>2001, p18</p>	<p>Problem text:</p> <p style="text-align: center;">CD Sales</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Type</th> <th>Year 1</th> <th>Year 2</th> <th>Year 3</th> </tr> </thead> <tbody> <tr> <td>Rock</td> <td>39.4%</td> <td>36.2%</td> <td>35.6%</td> </tr> <tr> <td>Pop</td> <td>17.8%</td> <td>16.5%</td> <td>15.5%</td> </tr> <tr> <td>Rap</td> <td>11.6%</td> <td>10.4%</td> <td>10.7%</td> </tr> <tr> <td>Country</td> <td>7.7%</td> <td>7.9%</td> <td>7.8%</td> </tr> <tr> <td>Other</td> <td>23.5%</td> <td>29.0%</td> <td>30.4%</td> </tr> <tr> <td>Total Sales (In thousands)</td> <td>\$1,732.10</td> <td>\$2,269.30</td> <td>\$2,639.80</td> </tr> </tbody> </table> <p>This table shows the distribution of CD sales of different types of music in a local store over a three-year period. Based on this information, from Year 1 to Year 3 the total dollar value of this store's sales of pop music.</p> <p>Breakdown using scaffold questions:</p> <ol style="list-style-type: none"> First, approximately how much money did pop music make in year one? <ol style="list-style-type: none"> To find out how much money pop music made in year one, you must multiply the total sales by the percent you are looking for, divided by 100. Total sales * (Percent/100) = individual sales. So in this case, you would multiply the total amount of sales during year one, 1732.10, by what number? So, now using the equation Total sales * (Percent/100) = individual sales. approximately how much money did pop music make in year one? Now, how much did pop music make in year three? Knowing the value of the sales of pop music in year one and year three, you can now answer the original question. From Year 1 to Year 3 the total dollar value of this store's sales of pop music <p>Breakdown using hints:</p>	Type	Year 1	Year 2	Year 3	Rock	39.4%	36.2%	35.6%	Pop	17.8%	16.5%	15.5%	Rap	11.6%	10.4%	10.7%	Country	7.7%	7.9%	7.8%	Other	23.5%	29.0%	30.4%	Total Sales (In thousands)	\$1,732.10	\$2,269.30	\$2,639.80
Type	Year 1	Year 2	Year 3																										
Rock	39.4%	36.2%	35.6%																										
Pop	17.8%	16.5%	15.5%																										
Rap	11.6%	10.4%	10.7%																										
Country	7.7%	7.9%	7.8%																										
Other	23.5%	29.0%	30.4%																										
Total Sales (In thousands)	\$1,732.10	\$2,269.30	\$2,639.80																										

2005, p16

Problem text:

On an airline, approximately 10% of the airline passengers who are booked for a flight do not show up for the flight. The airline has booked 160 passengers for a flight with maximum seating of 135. How many of the 160 passengers booked for this flight will not have a seat, assuming 10% of the booked passengers do not show up?

Breakdown using scaffold questions:

1. First, how many people actually made the flight?
 - 1 Subtract 10% of 160 from 160.
 - 2 to get 10% of 160, multiply by .1 (any easy way to do this for 10% is to move the decimal place over one to the left.)
2. How many of these people are left over when all the seats are taken?
 - Subtract the 135 seats from the people who made it.

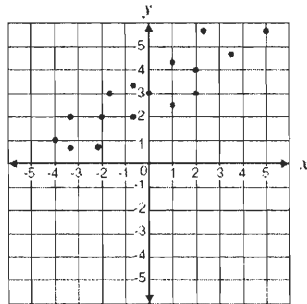
Breakdown using hints:

In response to specific problems:

- 1 Don't count the people who missed the flight as missing seats.
- 2 Not enough people miss the flight for everyone who shows up to get one.
- 3 No, this is the number who don't show up.

2002, p10

Problem text:



Use the graph below to answer the question. Which of the following equations best represents the data in the graph?

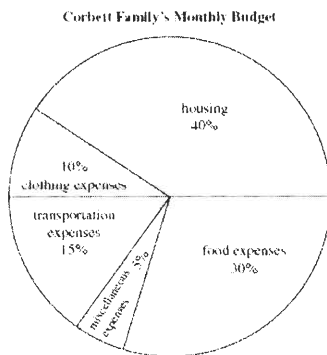
Breakdown using scaffold questions:

1. Approximately where does the set of points cross the y axis?
2. Approximately what is the slope of the points?
 - Take two points on opposite ends of the graph, find the y distance between the two points, and put that over the x distance between the points.
3. Now, put the equation in $y=mx+b$ form.

Breakdown using hints:

2002, p25

Problem text:



The circle graph above shows the Corbett family's monthly budget. The Corbett family has a total monthly income of \$2,000. Mr. Corbett received a \$100 per month raise. He increased the transportation expense portion of the monthly budget by \$100. To the nearest percent, what portion of the Corbetts' income is now being spent on transportation expenses?

Breakdown using scaffold questions:

1. First, how much was the Corbett family spending on transportation expenses?
 - 1 Multiply the original income by 0.15.
2. To the nearest percent, what portion of the Corbetts' income is being spent on transportation now?
 - Add 100 dollars to the number you came up with in the last problem to get the current transportation expenses.
 - Divide the current transportation expenses by the new income.

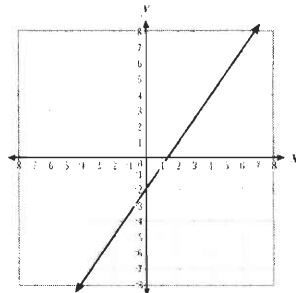
Breakdown using hints:

In response to specific answers:

- 1 Remember to add the \$100 to the expenses.
- 2 Remember to add the \$100 to the income.

2002r, p9

Problem text:

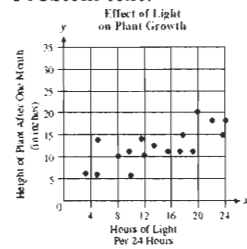


Which of the following best represents the equation of the line shown on the graph below?

Breakdown using scaffold questions:

1. There are two things to determine in finding the equation of a line, it's slop and it's y intercept. First, where does it cross the y axis?
2. ???MISSING???
 - Slope is determined by the change in X per the change in Y.
 - Start by finding to places on the line where you know the X and the Y positions, now, subtract the X on the left from the X on the right and do the same with the Ys, now divide the difference of the Xs by the difference of the Ys.
3. Now, which one is the correct equation?

Breakdown using hints:

Problem text:

Jenny studied the effect of light on plant growth. She graphed a scatter plot to represent her data.

Which of the following best represents the equation for the line of best fit for the data shown?

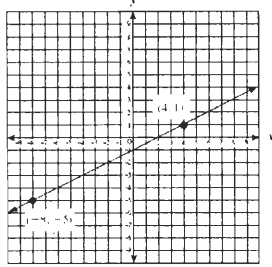
Breakdown using scaffold questions:

1. Look at the image above. The line in red is the line of "best fit" for the line shown. Can you figure out what its slope is?
 - The line is going up. So the slope must be positive.
 - Look at the incline. Does it seem that it goes upward faster than it goes to the right or vice versa?
 - It goes much faster to the right than upward.
 - So the slope is less than 1.
 - Looking at the choices in the first question, we see that the slope can be either 4 or 0.4, and we know that the slope is less than 1.
2. Ok. So what is the equation for the line of best fit for this data?
 - 1 The slope is positive. The slope is the coefficient of x .
 - 2 The slope is 0.4.
 - 3 So the answer is $y = 0.4x + 5$.

Breakdown using hints:

1. The line of best fit has this property: all the elements of the plot are somewhere in its vicinity.
2. All the elements in the plot are somewhere around the value 110.
3. The line of best fit is horizontal in this case.
4. The slope of a horizontal line is 0. Which equation shows a line with a slope of 0?
5. The answer is: $y = 110$.

Problem text:



Write an equation for the line drawn above.

Breakdown using scaffold questions:

- The equation you need to write is in the **slope/intercept** form. What does the slope/intercept form of a line look like?
 - The slope is m .
 - The y-intercept is b .
 - A straight line is linear, while a curve is quadratic.
 - The slope/intercept form of a straight line is $y = mx + b$.
- What is m , the slope of the line on the graph?
 - If you have a line that goes through two points, then its slope is the difference of the y-coordinates divided by the difference of the x-coordinates. This is also known as rise over run.
 - In this case, the line goes through $(-8, -5)$ and $(4, 1)$.
 - So the slope is $(1 - (-5))/(4 - (-8)) = (1 + 5)/(4 + 8)$
 - Which makes $m = 6/12$ or $1/2$.
- Good, now the equation for the line looks like $y = (1/2)x + b$. What is the value of b , the y-intercept?
 - The line goes through the point $(4, 1)$. That means if you replace x with 4 and y with 1 in the equation, you can easily find b .
 - If you follow the above hint, you'll get: $1 = (1/2) * 4 + b$.
 - Move all the terms that do not contain b on the left side and leave b on the right.
 - You now have: $1 - (1/2) * 4 = b$.
 - Which makes $b = -1$.
- Good. So what is an equation for the line in the graph?
 - The slope/intercept form of the equation is $y = mx + b$
 - The slope $m = 1/2$
 - The y-intercept $b = -1$
 - The answer is $y = (1/2)x - 1$

Breakdown using hints:

- So the answer is $y = 0.4x + 5$.
- The slope of a line that goes through two points is calculated as the ratio of the difference of the y-coordinates to the difference to the x-coordinates of those two points.
- If we choose $(4, 1)$ and $(-8, -5)$, then the slope is $1/2$.
- Since you already know the slope, you can find the intercept by substituting 4 for x, 1 for y, and $1/2$ for m in the slope/intercept form of the equation. You will get $1 = (1/2) * 4 + b$.
- Which makes b, the intercept, -1.
- The answer is $y = (1/2)x - 1$.

Appendix D

```
#!/usr/bin/perl
# Simple program to read and report on student scores on assignments
# from data file.
#
# Data file entry looks like:
# 2006-01-27 10:26:11.0084      2006-01-27 10:33:04.00647      26104  2004re37_exp  1.0
# start time                  end time                  userid  assignment      score

use strict;
use Time::ParseDate;

# Global variable to hold each students' records
my $students = {};

while (my $record = <STDIN>)
{
    # Discard garbage input lines.
    next if length($record) <= 5;

    # Variables to hold each entry in this input line.
    my $start_time;
    my $end_time;
    my $userid;
    my $assignment;
    my $score;

    # Parse input line for this assignment (for one particular student)
    ($start_time, $end_time, $userid, $assignment, $score) = split(/\t/, $record);

    # Add assignment to global student record.
    $students->{$userid}->{$assignment}->{'score'}           = $score+0;
    $students->{$userid}->{$assignment}->{'start_time'}     = parsedate($start_time);
    $students->{$userid}->{$assignment}->{'end_time'}       = parsedate($end_time);
}

# Students are required to complete the first, the last,
# and EITHER OF the two middle sets of the following problems:
my @required_pretest_assignments =
# ('pretest_triangle_2002r_40', 'pretest_triangle_2004_33');
# ('2004re37_exp', '2002r-10-11');
('linpretest-2005-10', 'linpretest-2005-35');

my @alternate_exp_assignments_1 =
# ('expl_scaffolding_2001_32', 'expl_scaff_2001item17', 'E1_2004r25_scaf',
'E1_2000t37_scaf');
# ('1999_14', '2001item18', 'E2_2005t16_scaf', '2002-g10-25');
('2002r-10-09', '2002-g10-10', '2003_LineBestFit', '2003_Retest_10g_18');

my @alternate_exp_assignments_2 =
# ('expl_noscaff_2001_32', '2001_17_gr10', 'E1_2004r25_hints', 'E1_200t37_hints');
# ('exp2_noscaff_1999_14', 'exp2_noscaff_2001_18', 'E2_2005t16_hints', '2002-g10-25c');
('2002_10_gr10', '2002r-10-9', '2003r_18-hints', '2003_39_gr10');

my @required_transfer_assignments =
# ('2002r-10-40', 'exp_2004_33', '2002r-10-26', 'T1_2004n40');
# ('exp2_trans_2004R_37', '2002r-10-11t', '2005rMarch3', '2005_37_gr10');
('2005-g10-10t', '2005-g10-35t', '2003r-g10-26', '2004r-g10-7');

my $exp=3;

#Exp 2 need to apply s/2002-g10-25/2002-g10-25c/ to RS-2 and apply s/2002r-10-11/2002r-10-11t/
to Transfer
#Exp 3 need to apply first two transfers

print
"\nUID\\", \"p1\\", \"p2\\", \"te0\\", \"e1\\", \"e2\\", \"e3\\", \"t0\\", \"t1\\", \"t2\\", \"t3\\", \"TimeTak
en\\", \"Condition\\", \"Exp\\n\";
```

```

foreach my $userid ( keys %{$students} )
{
    # boolean flag to indicate problem w/ curics.
    my $curics_error = 0;
    my $condition=0;

    my $assignments_count = scalar(keys( %{$students->{$userid}} ));

    if ( $assignments_count >= 10 &&
( $students->{$userid}->{$required_pretest_assignments[0]}->{'start_time'}<parsedate('200
6-01-20 10:20') ||
$students->{$userid}->{$required_pretest_assignments[0]}->{'start_time'}>parsedate('2006-
01-20 11:00'))
    {
        # Determine which problems this particular student did.
        my @problems;
        if ( exists $students->{$userid}->{@alternate_exp_assignments_1[0]} )
        {
            @problems =
            (
                @required_pretest_assignments,
                @alternate_exp_assignments_1,
                @required_transfer_assignments,
            );
            $condition = 1;
        }
        elsif ( exists $students->{$userid}->{@alternate_exp_assignments_2[0]} )
        {
            @problems =
            (
                @required_pretest_assignments,
                @alternate_exp_assignments_2,
                @required_transfer_assignments,
            );
            $condition = 2;
        }
        else
        {
            print "Error with curics, no scores will be shown. ($userid)\n";
            $curics_error = 1;
        }
        my $timeTaken = larger(
            $students->{$userid}->{$problems[2]}->{'end_time'},
            $students->{$userid}->{$problems[3]}->{'end_time'},
            $students->{$userid}->{$problems[4]}->{'end_time'},
            $students->{$userid}->{$problems[5]}->{'end_time'},
        ) -
        smaller(
            $students->{$userid}->{$problems[2]}->{'start_time'},
            $students->{$userid}->{$problems[3]}->{'start_time'},
            $students->{$userid}->{$problems[4]}->{'start_time'},
            $students->{$userid}->{$problems[5]}->{'start_time'});

        #print
        #           '1:', $students->{$userid}->{$problems[2]}->{'end_time'}, '
        #           '2:', $students->{$userid}->{$problems[3]}->{'end_time'}, '
        #           '3:', $students->{$userid}->{$problems[4]}->{'end_time'}, '
        #           '4:', $students->{$userid}->{$problems[5]}->{'end_time'}, "\n";

        if($timeTaken < 6000)
        {
            print "\"$userid\", \"";

            foreach my $problem (@problems)
            {
                print $students->{$userid}->{$problem}->{'score'};
            }
        }
    }
}

```


Acknowledgments

We would like to thank Cathi Luperchio at Worcester North High school, Leena Razzaq, and Pat Mckeon of Worcester Vocational High school.