

EDUCATION STUDY REPLICATION USING ASSISTMENT
The Use of Computer-Based Delivery Instead of Traditional Paper-and-Pencil

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

Matthew L. Crocker

Samuel L. Moniz

Ethan T. Truong

Date: March 5, 2010

Professor Neil T. Heffernan, Major Advisor

Cristina Heffernan, Co-Advisor

Abstract

In this replication study, we used the "Compared With What?" study by Bethany Rittle-Johnson and Jon R. Star as our model. The purpose of their study was to find what specific comparisons in math education should be emphasized in order to maximize learning and transfer. Here, we used the same material of assessment and intervention as the original model as students learned to solve equations by comparing (a) equivalent problems solved with the same solution method, (b) different problem types solved with the same solution method, and (c) equivalent problems with different solution methods. The primary difference between the two studies is our use of a computer-based delivery system ASSISTment, whereas the original study used traditional paper-and-pencil methods. The comparison results of this study are not statistically significant because the pair wise distributions of the three groups failed the t-test. Even though they may not be statistically significant, there was overall learning regardless of the comparison emphasis, and the data does suggest the same conclusions as the original study: comparing solution methods best support students' conceptual knowledge and procedural flexibility.

Table of Contents

Abstract	2
Table of Contents	3
List of Figures	4
List of Tables.....	5
Background	6
About ASSISTment.....	11
Method	20
Results	26
Discussion	32
Glossary of Terms	33
Bibliography.....	35
Appendix A: Assessment (Pre/Posttest).....	A1
Appendix B: ASSISTment Intervention	A13
Appendix C: Skills Built for ASSISTment	A86

List of Figures

Figure 1: "Compared With What?" Intervention Sample	9
Figure 2: "Compared With What?" Intervention Practice Problems Sample	12
Figure 3: Intervention Practice Problem Built With ASSISTment.....	13
Figure 4: "Compared With What?" Intervention Practice Problems, Solve Two Ways	14
Figure 5: Intervention Practice Problem Built With ASSISTment, Solve Two Ways.....	15
Figure 6: "Compared With What?" Intervention Sample	16
Figure 7: Intervention Practice Problem Built With ASSISTment.....	17
Figure 8: Equation Building Process.....	21
Figure 9: Compare Identical Worked Example	21
Figure 10: Compare Problem Types Worked Example	22
Figure 11: Compare Solution Methods Worked Example.....	22
Figure 12: Presentation of CI and CPT's Practice Problem.....	23
Figure 13: First Step in CSM Practice	23
Figure 14: Second Step in CSM Practice	23
Figure 15: Conceptual Knowledge by Condition	28
Figure 16: Procedural Knowledge by Condition.....	29
Figure 17: Flexibility Knowledge by Condition	30

List of Tables

Table 1: Grading Breakdown	25
Table 2: p-values for Compare Conditions.....	26
Table 3: Means and Standard Deviations for Assessment Scores, as Percents	27
Table 4: Means and Standard Deviations for High Skill and Low Skill Students, as Percents	30
Table 5: Correlation of Assessment Questions	31

Background

Existing Research on Comparison

Bethany Rittle-Johnson and Jon Star found three key findings in experimental studies on comparison: “two examples are better than one”, “two examples presented together are better than two examples presented separately”, and “instructional support augments the benefits of comparison” (Rittle-Johnson, 2009, p529).

Other relevant findings and experiments are listed here.

In Edward Silver’s (et al.) 2005 article, “Moving from rhetoric to praxis: Issues faced by teachers in having students consider multiple solutions for problems in the mathematics classroom,” Silver worked with teachers in the “BI:FOCAL project” to understand why the practice of comparison is rarely used in United States classrooms. At the conclusion of the program, teachers began “using multiple solutions as a complex, nuanced aspect of instructional practice,” finding this to be a more effective approach to teaching than simply identifying comparisons (Silver, 2005).

Mary Gick and Keith Holyoak’s 1983 article “Schema Induction and Analogical Transfer” investigates how analogies are noticed and applied to problem solving. They found that only 20% of their test subjects were able to spontaneously identify a specific analogy in solving a problem without be explicitly told to do so. However, Gick and Holyoak believe that a shared semantic component between two problems will help individuals recognize analogies (Gick, 1983, p17).

Stephen Reed’s 1989 article “Constraints on the Abstraction of Solutions” discusses three experiments, one of which involved giving college algebra students two example problems and solutions with different features that are shown to be similar and asking them to solve a third problem with different features that could be solved the same way. The students did not recognize that the same solutions could be applied. This contrasts Gick and Holyoak’s idea that a shared semantic component will help individuals recognize analogies, because even when told of the similarities, the students could not apply the analogy to solve the third problem (Reed 1989).

Mary Gick and Katherine Patterson’s 1992 article “Do Contrasting Examples Facilitate Schema Acquisition and Analogical Transfer?” used an approach similar to Reed’s in that two examples with similar solutions were shown to test subjects, followed by a third example that must be analyzed using the same approach as the first two. The study showed that using a near miss contrast (in which the examples are very similar, but not exactly the same) is more effective than using a third example that is unrelated to the first two (Gick, 1992).

In the 2007 article, “Does Comparing Solution Methods Facilitate Conceptual and Procedural Knowledge? An Experimental Study on Learning to Solve Equations,” it was found that students in a comparison condition had greater procedural knowledge and flexibility (Rittle-Johnson, 2007, p571).

About “Compared With What?”

In 2009, Rittle-Johnson and Star conducted a study to evaluate “three types of comparison for supporting middle school students’ conceptual and procedural knowledge of equation solving” (Rittle-Johnson, 2009). The study was called “Compared With What? The Effects of Different Comparisons on Conceptual Knowledge and Procedural Flexibility for Equation Solving.”

Three Modes of Comparison

Rittle-Johnson and Star defined three modes of comparison, which are abbreviated here as Compare Identical (CI), Compare Problem Types (CPT), and Compare Solution Methods (CSM). The "Compared With What?" study served to determine which of these three modes of comparison is the most efficient. Our study is a reproduction of this objective using a different medium.

CI is the comparison of “equivalent problems solved with the same solution method” (Rittle-Johnson, 2009, p532). When looking at two problems side by side, according to this mode, they will have the same problem type and solution method.

CPT is the comparison of “different problem types solved with the same solution method” (Rittle-Johnson, 2009, p532).

CSM is the comparison of “the same problem solved with two different solution methods” (Rittle-Johnson, 2009, p532).

Target Domain and Outcomes

Rittle-Johnson and Star selected linear equation solving as the focal point for their comparisons under study. They chose this mathematical concept because it is considered a fundamental skill, its methods and arguably the longest and most complex when first introduced, and because students are required to learn how to manipulate multiple rules instead of just one (Rittle-Johnson, 2009, p531).

Three types of linear equations were used in the “Compared With What?” study of the following basic structures:

Divide composite [$a(x + b) = c$]

Combine composite [$a(x + b) + d(x + b) = c$]

Subtract composite [$a(x + b) = d(x + b) + c$]

In the intervention/instruction portion of the experiment, students received two side-by-side solutions to each equation structure. The first, conventional method, involved applying the distributive property as the first step. The second, less conventional shortcut, involved treating $(x + b)$ as a single composite variable. The second method was used because it could “push children to understand important problem features and reflect on when different methods are more efficient” (Rittle-Johnson, 2009, p532).

The three target outcomes of “Compared With What?” were procedural knowledge, procedural flexibility, and conceptual knowledge. Procedural knowledge measures the ability to adhere to step-based problem solving. Procedural flexibility is the ability to understand multiple solutions to a problem and utilize them efficiently. Conceptual knowledge measures the ability to grasp the theory behind solutions (Rittle-Johnson, 2009, p532).

Rittle-Johnson and Star hypothesized based on their earlier 2007 study that CSM (Compare Solution Methods) would lead to the greatest improvement in all three of the target outcomes (Rittle-Johnson, 2009, p532). CSM demonstrates that there are two ways to solve the same problem. By seeing the same problem in two instances, the student will have a greater opportunity to recognize that they are connected than if the problems were similar or completely different.

Original Method

“Compared With What?” followed a pretest-intervention-posttest-retention test design. The intervention and all of the assessments were administered by paper and pencil. Students answered explanation prompts, solved practice problems, and received mini-lectures. Participating students already learned the distributive property, simplifying expressions, and solving one/two-step equations beforehand. They were assigned randomly in pairs to each of the three comparison types (Rittle-Johnson, 2009, p532-533).

The intervention portion of the experiment involved three different packets: one for each comparison type. Each student completed one packet—the one corresponding to his/her comparison type. The packet was split into three days. Each day included two pairs of side-by-side solution methods that ask for the types of operations being performed, followed by two to three follow up questions and four practice problems. A sample from the original intervention is pictured in Figure 1.

A. Compare Solution Methods

Nathan's Solution:	Patrick's Solution:
$5(y+1) = 3(y+1) + 8$ $5y + 5 = 3y + 3 + 8$ $5y + 5 = 3y + 11$ $2y + 5 = 11$ $2y = 6$ $y = 3$	$5(y+1) = 3(y+1) + 8$ $2(y+1) = 8$ $y+1 = 4$ $y = 3$
<i>Distribute _____</i> <i>Combine _____</i> <i>Subtract _____ on Both</i> <i>Subtract _____ on Both</i> <i>Divide _____ on Both</i>	<i>Subtract $3(y+1)$ on Both</i> <i>Divide _____ on Both</i> <i>Subtract _____ on Both</i>

- Describe 2 ways that these students' solutions are **different**.
- What must be true about an equation for Patrick's way to be easier than Nathan's way?

Figure 1: "Compared With What?" Intervention Sample

The assessment portion of the experiment was exactly the same for the pretest, posttest, and retention test. All three tests were identical. A modified version of the assessment test from Rittle-Johnson and Star (2007) was used to assess the three target outcomes. Conceptual knowledge was evaluated through verbal and nonverbal knowledge of algebraic concepts, procedural knowledge was gauged as the ability to solve equations, and procedural flexibility was assessed as the ability to use efficient solution methods and evaluated through flexible knowledge of solution methods (Rittle-Johnson, 2009, p534).

Original Procedure

The experiment took place during participating students' normal class times. It was split into five days.

The pretest took place on Day 1. It was timed for 40-50 minutes. Time limits were used to encourage students to use efficient solution methods (Rittle-Johnson, 2009, p534).

Day 2 was the first part of the intervention. It began with a ten minute introduction. The students attempted to solve $3(x+1) = 12$ on their own. Then, the instructor worked through the conventional solution with the class. Students worked on the packets in pairs (each has his/her own packet) and solved the practice problems individually before discussing them with their partners and checking their answers with an adult. Adults helped implement steps, but not choose steps or answer reflection questions (Rittle-Johnson, 2009, p534-535)

The intervention continued on Days 3-4. The same format was continued for each day, but an introductory lesson introduced variables on both sides of an equation on Day 3 and introduced fractional coefficients on Day 4. Students did not return to complete previously incomplete packets. An eight minute wrap-up lesson took place at the end of Day 4 that emphasized "There is more than one way to solve an equation," "Any way is okay if two sides of the equation are kept equal," and "Some ways of solving equations are easier than others." (Rittle-Johnson, 2009, p535)

The posttest took place on Day 5. It was identical to the pretest. An identical retention test was also given to the students two weeks later (Rittle-Johnson, 2009, p535).

Original Findings

Comparing solution methods was the most effective at enhancing conceptual knowledge and procedural flexibility. CSM, CPT, and CI were all equally effective at enhancing procedural knowledge (Rittle-Johnson, 2009, p541).

If a student was absent from two assessments, he/she was omitted from the results. If a student was absent for only one assessment, imputation was used instead (Rittle-Johnson, 2009, p535).

About ASSISTment

ASSISTment is a web service that provides a medium for educators to collaborate and interactively reach many more students than what would be possible without the Internet. Its primary function is building interactive tests and tutoring online with an interface that is improving each year. The makers of these tests can be teachers, researchers, or anyone who wishes to contribute to learning and education; teachers can then assign these test and tutoring modules to their classes and share them with their colleagues.

ASSISTment has two major features that complement its test-building abilities: tutoring and logging events. The tutoring system built into the testing infrastructure presents two primary methods of tutoring: giving hints and offering interactive sub-problems for the student to solve in order to understand the original problem. The other feature, the real-time event logging, allows an educator to see the students' progress through their tests in real-time. Some of the details logged include the correctness of the students' answers, the students' answers themselves, and the time the student spends on a problem or tutoring session before moving on to the next problem or tutoring session. This allows the educator running the test myriad opportunities, including identifying students who are abusing the tutoring system, finding common errors across the student body, and easily importing the already digital data to a spreadsheet to make further inferences.

ASSISTment and the “Compared With What” Study

ASSISTment has a lot to offer to the learning sciences. One of these explorations into ASSISTment's potential is the replication of the “Compared With What” study conducted by Rittle B. Johnson. Here we will use many of ASSISTment's features in order to replicate this study.

We utilized the built-in tutoring system to make sure that students do not get stuck on a single problem along the way, and ASSISTment provides logs of student uses of the tutoring system for us to check.

The instant feedback system allows students to know if they reached the correct answer; this is especially relevant when the study is not conducting tests but rather using worked examples for learning purposes.

Assessment's ability to log data proved invaluable to infer the levels of success in the study's multifaceted areas. Our collection and analysis of this data will be discussed in the "Results" section of this paper.

Procedural Differences

While the students' assessments were replicated just as the Rittle-Johnson and Star did it, there were a few alterations that our replication study made to the intervention in order to fit within the ASSISTment operation scheme and the participating teachers' established curriculum.

Material Changes

One of these changes is a condensed version of solving equations. In the original study, it required students to show their work (see Figure 2).

SOLVE THESE PROBLEMS BY YOURSELF. Be sure to show your work. Then, check your answers with a teacher.

$$\frac{1}{6}(x + 1) = 1$$

$$\frac{1}{9}(x + 3) = 1$$

$$4(x - 3) = 2(x - 3) + 10$$

$$4(f + 5) + 18 = 10(f + 5)$$

Figure 2: "Compared With What?" Intervention Practice Problems Sample

However, since these problems in the intervention were not graded, we allowed the student to simply enter the answer (see Figure 3).

The screenshot shows a web browser window titled "Assistment - Previewing Content - Google Chrome" with the URL <http://assistment.org/build/preview/assistment/77069>. The page features the ASSISTment logo and a blue banner stating "You are previewing content." and "72466 - Rittle-Johnson CPT, Day 1-3a (#77069)". The main content area contains the instruction "Solve this equation for h:" followed by the equation $12 = 3(h - 3)$. A link for "Comment on this question" is visible on the right. Below the equation is a blue button labeled "Show me hint 1 of 2". Underneath is the prompt "Type your answer below (mathematical expression):" and a text input field. At the bottom of the problem area is a blue button labeled "Submit Answer".

Figure 3: Intervention Practice Problem Built With ASSISTment

The primary reason for this alteration was our inability to automate grading the students' work towards the solution, and we would lose one of the advantages that ASSISTment offers: instant feedback to assist student learning. We made a similar change to problems asking the students to solve problems using two different methods. Again, since this was part of the intervention, these solutions were not graded (see Figure 4).

SOLVE THESE PROBLEMS BY YOURSELF. You will solve each problem **two ways**. Be sure to show your work. Turn your answers with a teacher.

	$\frac{1}{6}(x + 1) = 1$	
Way 1		Way 2
$\frac{1}{6}(x + 1) = 1$		$\frac{1}{6}(x + 1) = 1$
$4(x - 3) = 2(x - 3) + 10$		
Way 1		Way 2
$4(x - 3) = 2(x - 3) + 10$		$4(x - 3) = 2(x - 3) + 10$

Figure 4: "Compared With What?" Intervention Practice Problems, Solve Two Ways

In our replication study, we asked the student to choose the two possible equations out of a selection of equations that could be the next step in finding the solution, and then solve for the unknown (see Figure 5).

$15 = 5(x - 2)$

Select the two options below that could be the next step in solving this equation. [Comment on this question](#)

Show me hint 1 of 2

Select all that apply:

$15 = 5x - 10$

$3 = 5(x - 2)$

$3 = 5x - 10$

$3 = x - 2$

$15 = 5x - 2$

Submit Answer

✓ Correct!

Solve for x

$15 = 5(x - 2)$

[Comment on this question](#)

Show me hint 1 of 2

Type your answer below (mathematical expression):

Submit Answer

Figure 5: Intervention Practice Problem Built With ASSISTment, Solve Two Ways

Again, this allowed us to offer instant feedback to the students as they progress through these learning problems.

The last material change to the intervention was to the group-work problems asking students to solve an equation using the same method shown in the worked example.

Jessica's Solution:	Mary's Solution:
$3(t-1) + 3(t-1) = 30$ $3t-3 + 3t-3 = 30$ $6t-6 = 30$ $6t = 36$ $t = 6$	$3(t-1) + 3(t-1) = 30$ $6(t-1) = 30$ $t-1 = 5$ $t = 6$
<p style="text-align: right;"><i>Distribute _____</i></p> <p style="text-align: right;"><i>Combine 3t's and 3's</i></p> <p style="text-align: right;"><i>Add _____ on Both</i></p> <p style="text-align: right;"><i>Divide _____ on Both</i></p>	<p style="text-align: right;"><i>Combine (t-1)'s</i></p> <p style="text-align: right;"><i>Divide _____ on Both</i></p> <p style="text-align: right;"><i>Add _____ on Both</i></p>

GPI: Working with your partner, solve the following equation using Mary's way.

$$6(x + 4) + 5(x + 4) = 22$$

Figure 6: "Compared With What?" Intervention Sample

Instead, we asked the student to individually select the equation that a particular solution method would produce next, and then solve (see Figure 7).

<p>Eric's Solution</p> $4(y + 4) = 2(y + 4) + 2$ $4y + 16 = 2y + 8 + 2$ $4y + 16 = 2y + 10$ $2y + 16 = 10$ $2y = -6$ $y = -3$	<p>Abby's Solution</p> $4(y + 4) = 2(y + 4) + 2$ $2(y + 4) = 2$ $y + 4 = 1$ $y = -3$
<p>Distribute <u>(1)</u></p> <p>Combine <u>(2)</u></p> <p>Subtract <u>(3)</u> on both</p> <p>Subtract <u>(4)</u> on both</p> <p>Divide <u>(5)</u> on both</p>	<p>Subtract <u>2(y + 4)</u> on Both</p> <p>Divide <u>(6)</u> on both</p> <p>Subtract <u>(7)</u> on both</p>

What is the first step to solving the following equation using Abby's way:

$$5(x + 2) = 3(x + 2) + 16$$

[Comment on this question](#)

Select one:

- $5x + 10 = 3x + 6 + 16$
- $5x + 10 = 3x + 22$
- $5(x + 2) - 16 = 3(x + 2)$
- $2(x + 2) = 16$
- $5(x + 2) = 16$

[Submit Answer](#)

Correct!

Solve for x

$$5(x + 2) = 3(x + 2) + 16$$

[Comment on this question](#)

[Show me hint 1 of 2](#)

Type your answer below (mathematical expression):

[Submit Answer](#)

Figure 7: Intervention Practice Problem Built With ASSISTment

Non-Material Changes

Students were drawn from multiple schools like the original study had done, but not all the classes consisted of regular and advanced mathematics classes; our sample included students in remedial classes. The time frame of our replication also differed from the original study in order to incorporate the needs of the participating teachers and their planned curriculum. Whereas the original study conducted all of their testing and intervention in five consecutive days, we allowed the participating teachers to start the study in the middle of week and continue after the weekend. We preserved the amount of work that was done each day by the original study.

Computer-Based and Paper-Pencil Testing

In some testing applications, computer-based delivery like that employed by ASSISTment is gaining popularity. It contains many potential advantages, including innovative interfaces afforded by the technology, reduced cost of production, administration, delivery, and scoring, and more that are described in detail below. One of the primary concerns that may arise is the comparability of computer-based testing and paper-pencil-test. Pearson (2009) summarized studies and study reviews on this topic, including Mazzeo and Harvey's review (1988) of some 30 studies with mixed evidence, Mead and Drasgow's study (1993), which reached similar conclusions, and Kim's tests (1999), which concluded that the average scores were comparable. Although Pearson concludes that incomparability will diminish over time, especially as computer unfamiliarity decreases, it cautions that alternative test taking methods should not be taken for granted.

+Scalability

One of the primary benefits of using ASSISTment and web technologies in general is scalability, that is, the ability to scale up in order to handle explosive growth, whether that growth comes from an increase in users, features, or a combination of both. Whereas the intervention which the "Compared With What" study is physically limited to students to whom researchers may send This is an especially important attribute to our study as we seek to increase the number of participants. This in turn will increase the amount of available data across a range students from varying locations, schools, and teachers, making our data all the more potent. With this increased amount of information to analyze, there will be opportunities to study specific aspects in greater detail which will entail more updates and features, updates and features for which ASSISTment can provide the tools.

+Availability

Another benefit of the transition to digitizing test taking and tutoring is the ease in which these digital resources are made available. Now that we have generated the program, any student or educator with internet access can take part in it. ASSISTment records the results

in real time, and we can grant anyone we deem appropriate the privileges required to see this data. This feature can be particularly appropriate for other researchers and educators who wish to run their own analysis on the data or point out anomalies to the education community. A paper and pencil methodology would make this sort of delivery cumbersome, and the results would have to be digitized to make it as readily available to the general public.

+Ease of collecting and manipulating digital results

Digital data is not only easy to collect (making it potentially very large) but also easy to analyze (despite its potential size), especially when compared to the traditional means with paper and pencil. ASSISTment creates this opportunity that has been previously unavailable, or at least prohibitively expensive before the movement in education towards digital technologies. This allows us to skip the step of manual data entry which other studies using paper and pencil must incorporate.

+Ease of making mass changes instantly

This feature ties in with our previous mention of scalability. Since there is only one instance of the study, if it requires an update, we can implement this update once in order for it to propagate to all of the study's participants. These updates can come from a need to fix an error or build a variation of the program in order to study a particular point of interest more closely. Incorporating new changes without this technology would require a complete re-print and re-delivery of all hard-copies: an intractable cost as studies become larger.

+Immediate feedback

Use of a personal computer also comes with the added advantage of more individual (albeit computerized) feedback to the participant. Without this technology, the only means to mimic this feature is to would be to have a 1:1 ratio of educator to student. This ability allowed us to circumvent the need Rittle-Johnson had of bringing a researcher to every intervention testing the students had to check their work as they progressed.

-Less Personal Interaction

There may be concerns that this technology is too impersonal and may be inferior to a human educator. It is our opinion that this technology is not meant to replace human teachers but rather serve as a tool for teachers to use. Any educational tool, no matter how powerful, should not be stretched beyond its means. This tool, however, can not only serve students but also identify those who do need more individual, personal attention, helping both the student to learn and the teacher to educate. This reduced personal interaction is merely an option; if a human educator is required, this system does not negate the possibility of bringing to enhance the students' education further.

Method

Design

We used the same design as the original study, a pretest-intervention-posttest design (Rittle-Johnson, 2009). We did not include a retention test. For the intervention, each student was randomly assigned to one of the three conditions. During the intervention each student worked through our problem set, answering questions about worked examples and solving equations in a manner related to their condition.

Materials

As mentioned previously, we had two goals with this study, our first goal was to replicate the study by Star and Riddle-Johnson and our second goal was to show a method that this type of study could be successfully performed using the ASSISTment platform (Rittle-Johnson, 2009). For both of these goals, our first task was materials. We had copies of the original study materials so over a period of a few weeks we built thirty-six different variabilized templates with ASSISTment.

We used the same base problems as the original study, if the equation that the student would have to solve in the original study was $3x + 2 = 5$ our version of that problem would look similar. The differences are one of the advantages of the ASSISTment system, the ability to create variabilized templates. $3x + 2 = 5$ would be turned into $AV + B = C$ where all of those letters are variables, expressing a range of possible values. For example A could be any number between 2 and 10, V would be a letter, B would be a number from 2 to 10 and C would be a number from C to 10. With the variabilized template we could then automatically generate a large number of assessments from one template. When variabilizing we had to be careful with the ranges we allowed so that we did not end up with difficult fractions as part of the solutions. To do this we would determine which variables are entirely random and which depend on other values. For the current example, C would equal a random number from 1 to 4 multiplied by A plus B. This would force the result to come out to be a positive integer so long as A and C were integers. We were also careful to make sure that the numbers never got too big or too small. Frequently if the range of the variables included 1 or 0, or other values depending on the problem, the problem could be trivialized. On the other end, if the numbers got too big we risk the difficulty of the problem to be in the realm of arithmetic instead of equation solving. Other than these differences, we attempted to make each day on our ASSISTment based study progress similarly to the paper based Riddle Johnson study (Rittle-Johnson, 2009). See Figure 8 for a diagram of the equation building process.

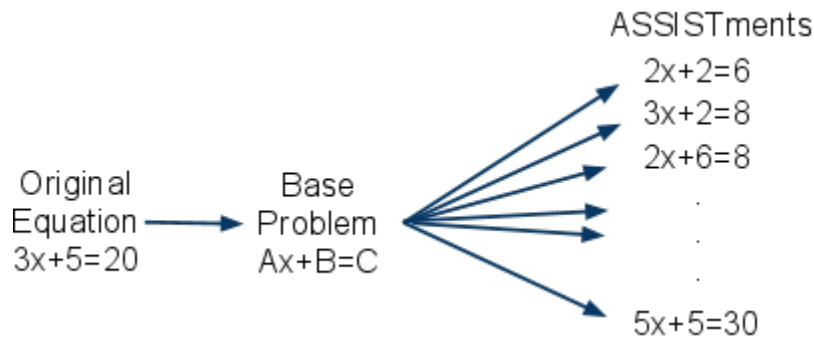


Figure 8: Equation Building Process

For all of our problems based in the ASSISTment system, we had to make sure that it was possible to progress. We did this by adding a pair of "hints" to each fill in response type. The first hint warned the student that there is no tutoring and the second gives the answer to the problem at hand. While this does allow for the student to just hint through the problem set we believed that preferable to a student being unable to progress at all.

For our pre and post tests we used the same exams that were used for Riddle Johnson 2009 and it is available in Appendix A (Rittle-Johnson, 2009). We decided to use the paper exams for a number of reasons. The first was it would make it simpler to compare our results to the original study. By using the same exams, we could use the same coding system. The other advantage to this is to see that the students are able to transfer what they used on the computer to a paper exam. The disadvantage to this is that it does not lend itself to our secondary goal, showing that a study of this design could be done entirely on the ASSISTment platform. However, our intervention was entirely digital.

Conditions

We tested the same three conditions that were tested in the original study. All three conditions focused on using comparisons to teach equation solving.

Compare Identical (CI) focused on comparing a problem to another problem that could be solved in the exact same manner. For this condition the student would be presented two problems side by side that share the same base problem. Because they share the same base problem they can be solved in the same manner. This is evident in Figure 9. The teaching method for this condition is to teach one solution method through repetition.

<p>Nathan's Solution</p> $5(h-2) = 10$ $5h-10 = 10 \quad \text{Distribute } \underline{5}$ $5h = 20 \quad \text{Add } \underline{(1)} \text{ on Both}$ $h = 4 \quad \text{Divide } \underline{(2)} \text{ on both}$	<p>Laura's Solution</p> $7(z-3) = 14$ $7z-21 = 14 \quad \text{Distribute } \underline{(3)}$ $7z = 35 \quad \text{Add } \underline{(4)} \text{ on both}$ $z = 5 \quad \text{Divide } \underline{(5)} \text{ on both}$
--	---

Choose what best answers (3)

Figure 9: Compare Identical Worked Example

Compare Problem Types (CPT) has different base problems that are related in such a manner that solutions are similar. The most common example of this is that one of the base problems has an additional initial step which turns it into a problem that shares the same base problem. Figure 10 shows an example of the worked example that a student put into Compare Problem Types would see. This method teaches by showing two different problems with two related solutions, it shows methods of reducing a problem that the student may not recognize so that it can be solved with a familiar method. For this condition, one of the problems is an expanded version of the other. The expanded problem after a few operations can be simplified so that it shares the same base problem, and can be solved with the same method as the other problem.

<p>Eric's Solution</p> $2(h - 2) + 2(h - 2) = 16$ $2h - 4 + 2h - 4 = 16 \text{ Distribute } 2\text{'s into } (h - 2)\text{'s}$ $4h - 8 = 16 \text{ Combine } 2h\text{'s and } 4\text{'s}$ $4h = 24 \text{ Add } \underline{(1)} \text{ on both}$ $h = 6 \text{ Divide } \underline{(2)} \text{ on both}$	<p>Alice's Solution</p> $2(h - 2) = 6$ $2h - 4 = 6 \text{ Distribute } 2 \text{ into } (h - 2)$ $2h = 10 \text{ Add } \underline{(3)} \text{ on both}$ $h = 5 \text{ Divide } \underline{(4)} \text{ on both}$
---	---

Fill in (3)

Figure 10: Compare Problem Types Worked Example

The final condition is Compare Solution Methods (CSM). This condition takes the same problem and solves it two different ways. This condition's goal is to show the student that many problems have multiple solution methods. Each worked example pair solves the same equation with two different methods (see Figure 11).

<p>Eric's Solution</p> $4(y - 7) = 12$ $4y - 28 = 12 \text{ Distribute } 4$ $4y = 60 \text{ Add } \underline{(1)} \text{ on both}$ $y = 15 \text{ Divide } \underline{(2)} \text{ on both}$	<p>Abby's Solution</p> $4(y - 7) = 12$ $y - 7 = 3 \text{ Divide } 4 \text{ on Both}$ $y = 10 \text{ Add } \underline{(3)} \text{ on both}$
--	---

Fill in (3)

Figure 11: Compare Solution Methods Worked Example

For each of these conditions we took the worksheets created for the original study and recreated them as ASSISTment variabilized templates. Each condition is broken down to three days and each day has the student do six different problems, based off of which condition they were randomly assigned to. The six problems are broken down into two comparison problems that are fairly lengthy and two pairs of practice problems.

The pairs of practice problems are related based off of which condition the student is in. For Compare Identical each problem in each pair of practice problems shares the same base problem. Compare Problem Types pairs are related so that one of the problems can be

solved so that it shares the same base problem as the other. Compare Identical and Compare Problem Types practice problem's presentation are simple, they just ask for the student to solve the equation and give the solution in a fill in box, as seen in Figure 12. Compare Solution Method however has a different layout, as the goal of the practice problem is for the student to solve the same equation using two different methods. Figure 13 shows how first the student is asked to choose all the valid first steps to solve the equation, they are then asked as seen in Figure 14. to give the solution to that equation, which looks identical to the Compare Identical and Compare Problem Types. The "pair" for Compare Solution Method is evident in the first problem where they are asked to find the two allowable first steps in solving the problem.

Solve this equation for d:

$$20 = 5(d - 5)$$

Figure 12: Presentation of CI and CPT's Practice Problem

12 = 3(x - 2)

Select the two options below that could be the next step in solving this equation.

[Cancel all responses](#)

Show me first 1 of 2

Select all that apply:

4 = 3(x - 2)

12 = 3x - 2

4 = 3x - 6

12 = 3x - 6

4 = x - 2

[Submit Answer](#)

Figure 13: First Step in CSM Practice

Solve for x

$$12 = 3(x - 2)$$

Figure 14: Second Step in CSM Practice

Participants

The participants for our study were students of teachers who have already used the ASSISTment system. We presented our ideas and plans for the study to a group of teachers that come to a regular meeting at Worcester Polytechnic Institute on the ASSISTment system. The presentation outlines the how the study can would be given and what the goals and outcomes of the study. The other method of gathering participants was having Cristina Heffernan speak with additional teachers to gather interest.

The students were drawn from five different schools with varying pre knowledge and learning abilities. We had 80 in total participate in the study, of which 50 we got usable data from. The reasons for exclusion were not taking one of the exams, or there were some students that did the wrong problem set for the intervention so they saw multiple conditions, or they did not finish a sufficient portion of the intervention, we excluded those that did not finish the worked examples in day three.

Procedure

To run the study we give any interested teachers copies of the pre and post tests. The teacher gives the pre test to the students and gives them a fixed time for each section of the test. To aid in this, we color coded the test so that it is easy to tell which section of the test each student is working on. Our foremost reason for timing the pre and post tests was that was the method that Riddle Johnson 2009 used.

The next step is in the computer lab. Over three days, not necessarily consecutive but within two weeks, the students do the problem set in ASSISTment. The problem set is set up to automatically assign each student to a condition. Each day the students will work on one day until completion. If a student does not complete a day they pick up where they left off the next day, or possibly are assigned to finish it as homework. These days were not necessarily consecutive; rather they were given at the teacher's convenience. Finally, after completing each of the three days the students take a posttest which is an identical test to the pretest.

Grading

To grade the pre and post tests, we split the exams into sections and had each section always graded by the same person. We based our coding system off of the system that was described by the original study (Rittle-Johnson, 2009) (see Table 1). See Appendix A for a copy of the pre and post test.

Table 1: Grading Breakdown

Part I	Solve 8 Equations	1 point for the correct answer for each
Part II	Solve two equations three ways each	For each equation, 1 point for solving it correctly two ways 2 points for solving it correctly three ways
Part III	Yes/No Questions, valid first steps to solve the given equations.	1 point for each correct answer
	Evaluate nonconventional methods	1 point for correctly identifying first step 2 points for identifying it is a good method, 1 point for identifying it is ok, but not a very good way 2 points if explanation notes efficiency or justifies why the solution is correct, 1 point if vague or it prefers a different method, 0 if wrong or blank 1 point for correct multiple choice
	Multiple choice questions	1 point for correct answer
	Equivalency without solving	1 point for correct multiple choice answer 1 point for explanation

For the intervention most of the problems were made so that they could be graded automatically by the ASSISTment system. This was done by making the problems multiple choice, fill in, or check all that apply. The other problems were considered "ungraded open response" by ASSISTment, which the system does show to us. However, we did not grade them. In general we did not grade the intervention; instead we maintained the data from the ASSISTment system so that we could examine it.

Results

Consequences of t-test

T-tests were used to determine the validity of our data (see Table 2). The tests were done in three pairs (two conditions to a pair) for the pretest, posttest, and gain.

Table 2: p-values for Compare Conditions

	PRETEST	POSTTEST	GAIN
CI and CPT	0.408	0.442	0.791
CI and CSM	0.714	0.719	0.862
CPT and CSM	0.202	0.173	0.611

According to these numbers, there is a large probability that any trends found between the different compare conditions will be the result of change—not significance. All of the p-values between the conditions exceeded the typical significance level of 5%

This means any conclusions made in this experiment cannot be made with confidence. All of the following trends are hypothetical, assuming any trends found are not caused by chance. The probability of this actually occurring for each scenario is 100% minus the corresponding p-value.

Pretest and Posttest Scores

The means and standard deviations of each condition for the pretest and posttest are presented in Table 3.

Table 3: Means and Standard Deviations for Assessment Scores, as Percents

	Pretest		Posttest		Effect Size
	Mean	SD	Mean	SD	
Conceptual					
CSM	23.08	13.73	40.66	20.90	1.281
CPT	17.26	16.84	28.57	16.85	0.672
CI	35.16	22.30	40.66	25.98	0.246
Total	23.43	18.87	34.86	21.04	0.606
Procedural					
CSM	23.08	20.31	23.08	16.81	0.000
CPT	21.88	17.77	25.52	18.61	0.205
CI	17.31	16.57	25.96	21.32	0.522
Total	21.00	17.94	25.00	18.56	0.223
Flexibility					
CSM	33.33	11.41	42.95	15.06	0.843
CPT	28.99	13.20	36.11	15.96	0.539
CI	30.77	10.96	38.46	19.48	0.702
Total	30.58	12.10	38.50	16.61	0.654

The range of the pretest results is 17.90%. Not all students entered the experiment with the same skill level with respect to solving equations. This will be taken into consideration in a proceeding section, but first the three conditions (CSM, CPT, and CI) will be compared according to the three target outcomes: conceptual, procedural, and flexibility knowledge.

Compare Conditions Applied to Target Outcomes (Posttest Data)

According to the posttest data, Figure 15 shows how the compare conditions fared with respect to Conceptual Knowledge.

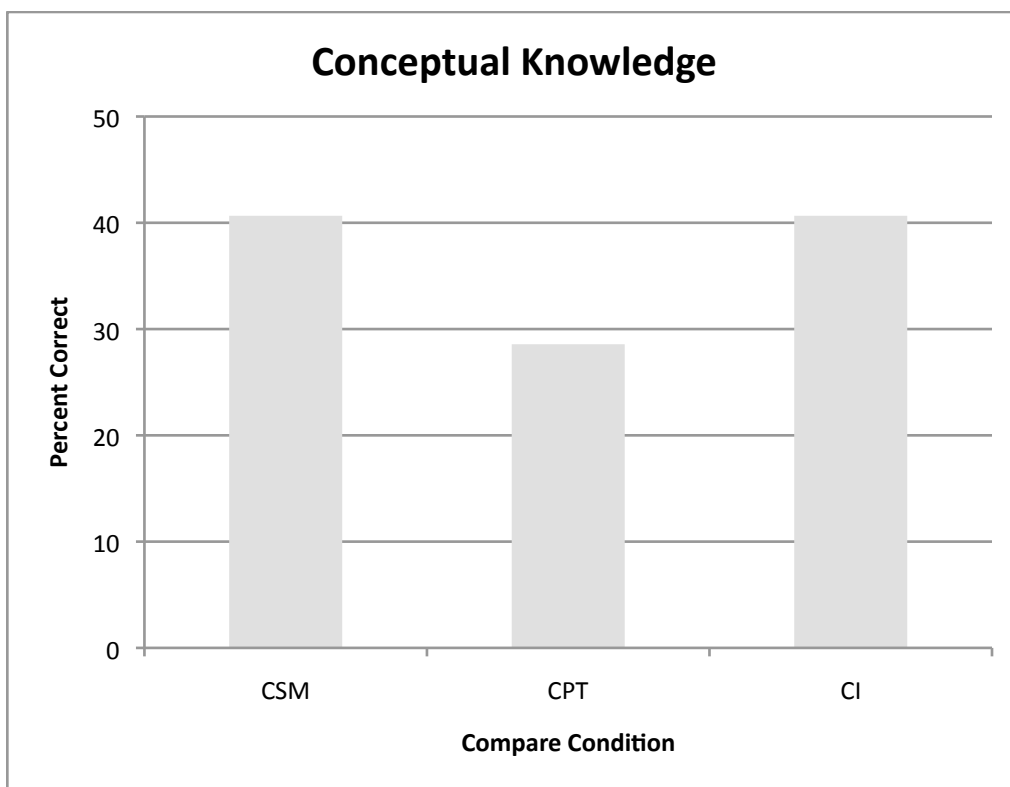


Figure 15: Conceptual Knowledge by Condition

CI and CSM had a greater effect toward mastering conceptual knowledge than CPT. Both seem to have had the same effect.

Figure 16 shows how the compare conditions fared according to Procedural Knowledge.

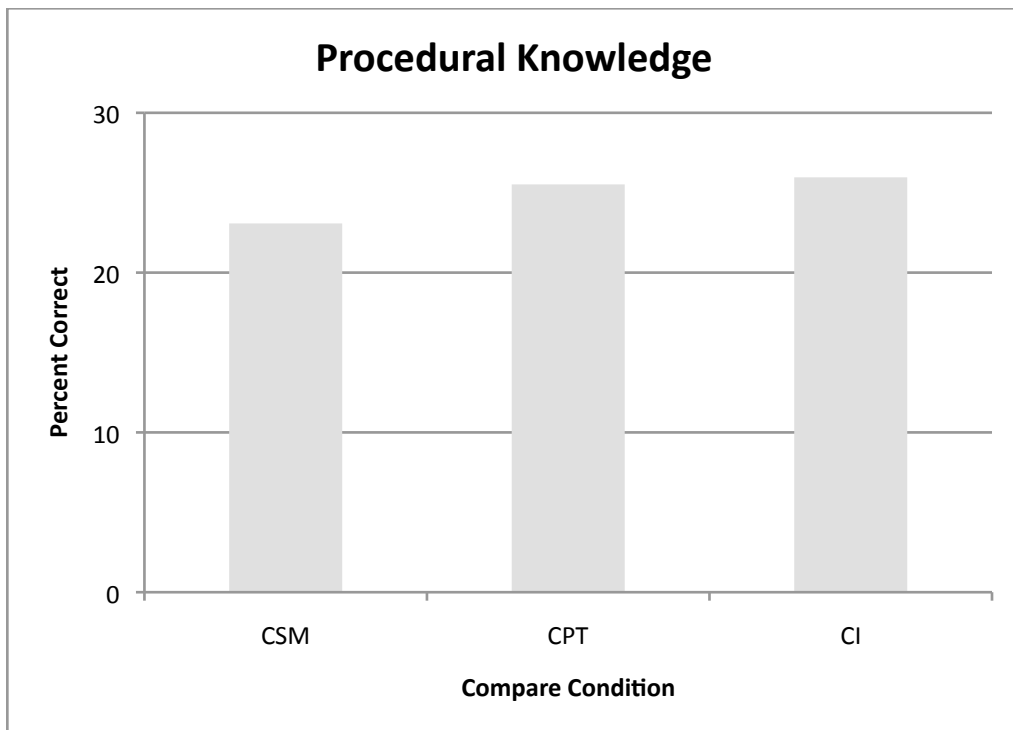


Figure 16: Procedural Knowledge by Condition

All three compare conditions had a relatively close effect on procedural knowledge. CI produced the greatest scores, but only by a small amount more than CPT and CSM.

Figure R3 shows how the compare conditions fared with respect to Flexibility Knowledge.

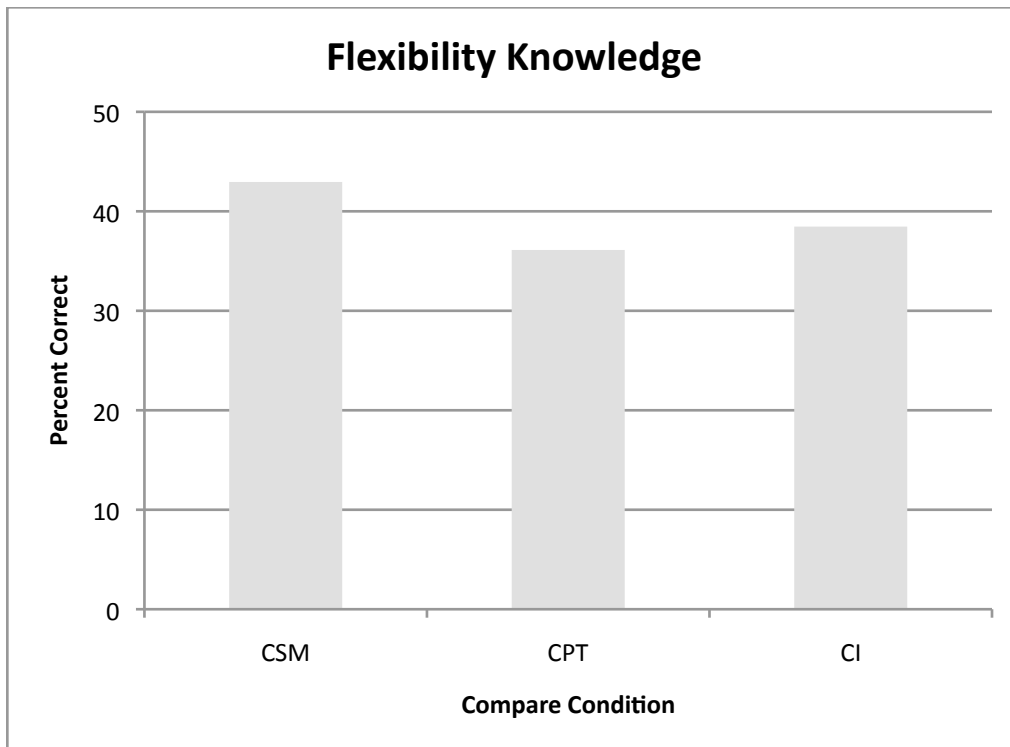


Figure 17: Flexibility Knowledge by Condition

CSM produced the greatest scores for flexibility knowledge, not much greater than the other two conditions.

Compare Conditions Applied to Pretest Knowledge

Based on their pretest scores, students were divided into two groups: high skill (score > median) and low skill (score < median). Table 4 shows the gain (from pretest to posttest) of high skill and low skill students according to each compare condition.

Table 4: Means and Standard Deviations for High Skill and Low Skill Students, as Percents

	Gain (High Skill)			Gain (Low Skill)		
	Mean	SD	Effect Size	Mean	SD	Effect Size
CSM	7.69	15.17	0.51	10.68	7.19	1.49
CPT	7.98	10.61	0.75	9.86	9.15	1.08
CI	13.25	16.77	0.79	4.03	7.69	0.52
Total	9.32	13.61	0.69	8.48	8.42	1.01

The highest gain was attributed to high skill students in the CI condition. The lowest gain was attributed to low skill students in the CI condition. Gains for CSM and CPT were higher for low skill students than for high skill students.

Correlation Analysis of Assessment Questions

Table 5 shows how well different assessment problems correlated with effect sizes throughout the study.

Table 5: Correlation of Assessment Questions

problem id	target outcome	correlation	problem id	target outcome	correlation
6	procedural	-0.120	10distribute	flexibility	0.189
7a	flexibility	-0.099	11a	flexibility	0.195
9combine	flexibility	-0.084	4	procedural	0.223
8a	flexibility	-0.052	5	procedural	0.277
18explain	conceptual	0.008	10subtract	flexibility	0.297
2	procedural	0.010	17	conceptual	0.305
10divide	flexibility	0.015	14	conceptual	0.309
18	conceptual	0.056	10combine	flexibility	0.328
3	procedural	0.085	11c	flexibility	0.349
1	procedural	0.088	9subtract	flexibility	0.365
16	conceptual	0.120	11b	flexibility	0.400
12a	flexibility	0.125	11bexplain	flexibility	0.441
8	procedural	0.126	12c	flexibility	0.453
9distribute	flexibility	0.140	9divide	flexibility	0.478
13	conceptual	0.183	12b	flexibility	0.551
15	conceptual	0.187	12bexplain	flexibility	0.614

Several problems assessing flexibility knowledge have the highest correlation values. Eight of the ten highest values are for problems 11 and 12. These problems asked students to assess the first steps of sample problems. Problems 7a and 8a (also assessments of flexibility knowledge) have very low correlation values. These problems asked students to solve equations more than one way.

Discussion

Improvements to Future Studies

This study did not give as useful results as we were hoping for. Unfortunately our data did not allow us to draw any conclusions about differences between the conditions. There are several possible reasons for this, our small number of participants, and external variables that we did not control during the study.

For the original study, Riddle-Johnson and Star 2009, they had 162 participants. These participants were drawn from nine different classes taught by five different teachers. We only had 50 participants after our previously discussed exclusion criteria. This is likely the main reason for our results. We would suggest to anyone else attempting to replicate this study to involve more students. We had 50 usable participants; to do this study effectively we should have 200 to 300 participants.

Another issue could be that we did not script the five days. Because we allowed the teachers to give the study as was convenient for them, not necessarily over consecutive days we introduces an additional external variable. This is because between the pretest and the posttest we could not control any math lessons that may be given by the teacher. Because those lessons would be split by teacher and not condition, it would reduce the spread for a class making it harder to tell if any gain was from our intervention or those lessons. While the ideal solution to this would be to script the days, due to the time commitment for the study and because most of the teachers do not have daily access to the computers necessary for our intervention, using scripted days and enforcing our days to be consecutive may have caused it to be impossible to get teachers to commit to the study.

The other uncontrolled variable was the prior knowledge of the students coming into the study. We accounted for the students' knowledge at the beginning of the study with the pretest. That test does not account for learning ability however. Our participants were drawn from average and remedial math classes, so there was a large range in prior knowledge. The original study drew its participants from average and high learning students.

What did we accomplish?

While we were unable to show a discernable difference between conditions, we did show that there was overall learning over the intervention. This shows that our intervention ASSISTment problem set has a value in its own right. While it is optimized for a study, it could easily be modified so that we have an end product that the ASSISTment team could give to teachers as a uniquely structured equation solving problem set.

Glossary of Terms

Assistment

An assistment is one problem that is given to the student to solve. It may or may not have tutoring associated with it.

ASSISTment

"The ASSISTment system is a web-based tutoring program for 4th to 10th grade mathematics. The word 'ASSISTment' blends tutoring 'assistance' with 'assessment' reporting to teachers. This gives teachers fine grained reporting on roughly 120 skills that the system tracks per grade level." ("About ASSISTments", TeacherWIKI)

Assistment Template

An assistment template is an assistment as it is being built by an educator.

Base Problem

To get a base problem we look at a normal math problem, e.g. $3x + 5 = 4$ and variabilize it, so that we retain the method of solving it while using unique numbers every time a student sees it. $3x + 5 = 4$ would change to $Ax + B = C$ for which A, B and C we could specify a range of numbers.

Check all that apply

A method for the student to provide the solution to an assistment. For this method, the student is given several possible answers and has to choose all answers that correctly answer the question.

CI

Compare Identical. One of our conditions. This one compares solving two similar equations (they share a base problem) using the same method.

Condition

Our study had three conditions that we tested to see which was best.

CPT

Compare Problem Types. One of our conditions. This one compares solving two different but related equations. One of the equations can be reduced so that it shares a base problem with the other and from that point they can be solved the same way.

CSM

Compare Solution Methods. One of our conditions. This one compares two different solutions to the same problem.

Fill In

A method for the student to provide the solution to an assistment. This answer method has a fill in box which the student can enter a value to and it is compared with the correct answer to see if the student is correct.

Multiple Choice

A method for the student to provide the solution to an assistment. For this method, the student is given several possible answers and has to choose one.

Problem Set

A collection of assistments that can be assigned to students. Problem sets can be built with varying structures so that students do not necessarily have identical experiences from the same problem set.

Ungraded open response

A method for the student to provide the solution to an assistment. This method is not graded by the system; it is only recorded for the teacher to look at.

Tutoring/Hints

Some assistments have tutoring associated with them. One of the possible types of tutoring are hints, which will walk the student through how to solve the problem.

Variabilized Assistment Template

A variabilized assistment template is an assistment template that in place of some numbers variables are used so that one template can generate a large number of assistments.

Bibliography

"About ASSISTments". WPI. 3/2/2010

<http://teacherwiki.assistment.org/wiki/index.php/About_ASSISTments>.

Gick, Mary L., and Keith J. Holyoak. Schema Induction and Analogical Transfer. Rep. 1983. University of Michigan. 2010

<http://reasoninglab.psych.ucla.edu/KH%20pdfs/Gick_Holyoak%281983%29_SchemaInduction.pdf>.

Gick, M. L., & Paterson, K. (1992). Do contrasting examples facilitate schema acquisition and analogical transfer?. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 46, 539-550. doi: 10.1037/h0084333

Hong Wang, Chingwei David Shin, Pearson. (2009). Computer-Based & Paper-Pencil Test Comparability Studies. PearsonAssessments Test, Measurement & Research Services Bulletin, November 2009, Issue 9.

Mazzeo, J., & Harvey, A. L. (1988). The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature. (College Board Report 88-8). New York: College Entrance Examination Board.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449-458.

Reed, S. K. (1989). Constraints on the abstraction of solutions. *Journal of Educational Psychology*, 81, 532-540. doi: 10.1037/0022-0663.81.4.532

Rittle-Johnson, Bethany, and Jon R. Star. Journal of Educational Psychology 99 (2007).

Rittle-Johnson, Bethany, and Jon R. Star. Journal of Educational Psychology 101 (2009).

Silver, Edward A., Hala Ghouseini, Dana Gosen, Charalambos Charalambous, and Beatriz T. Font Strawhun. Moving from rhetoric to praxis: Issues faced by teachers in having students consider multiple solutions for problems in the mathematics classroom. Rep. 3 Oct. 2005. University of Michigan. 2010

<[http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6W5B-](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6W5B-4H7SYHP-)

[2&_user=74021&_coverDate=12%2F31%2F2005&_rdoc=1&_fmt=high&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1219473728&_rerunOrigin=google&_acct=C000005878&_version=1&_urlVersion=0&_userid=74021&md5=6e3ab2e4a226795f68e4ed3f7c8de21b](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6W5B-4H7SYHP-2&_user=74021&_coverDate=12%2F31%2F2005&_rdoc=1&_fmt=high&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1219473728&_rerunOrigin=google&_acct=C000005878&_version=1&_urlVersion=0&_userid=74021&md5=6e3ab2e4a226795f68e4ed3f7c8de21b)>.