Loehr, A. M., Durkin, K., Rittle-Johnson, B., Star, J. R. (2019, April). Impact of comparison and explanation of multiple strategies on learning and flexibility in algebra. Paper presented at the American Educational Research Association (AERA), Toronto, Canada.

Impact of Comparison and Explanation of Multiple Strategies on Learning and Flexibility in Algebra Classrooms

Objectives

Current evidence suggests multiple strategies, comparison and explanation are all effective ways to support student learning (e.g., Alfieri, Nokes-Malach, & Schunn, 2013; Hodds, Alcock, & Inglis, 2014; Woodward et al., 2012). Thus, we developed a supplemental algebra curriculum comprised of worked example pairs and explanation prompts to increase comparison and interactive discussions in algebra classrooms. To measure the efficacy of our approach, we compare student learning in classrooms that used our teacher-led, supplemental curriculum with business-as-usual classrooms. We also explore how our supplemental curriculum impacted teachers' instructional practices that might support student learning.

Theoretical Framework

Recent theories of algebra learning stress the importance of supporting conceptual knowledge, or knowledge of abstract, general principles (e.g., Byrnes & Wasik, 1991; Kieran, 1992). But, competency in algebra requires both conceptual and procedural knowledge, or knowledge of mathematical strategies (e.g., Baroody, Feil, & Johnson, 2007; Byrnes & Wasik, 1991). Working with symbolic strategies is essential in algebra learning. Furthermore, existing theories do not focus on procedural flexibility, or knowing multiple strategies for solving a problem and selecting the most appropriate strategy for a given problem (Star & Rittle-Johnson, 2008; Woodward et al., 2012). Learners who develop procedural flexibility are more likely to use or adapt existing strategies when faced with unfamiliar problems and to have greater conceptual knowledge (Blöte, Van der Burg, & Klein, 2001; Hiebert et al., 1996). We argue that existing theories of algebra learning are incomplete and have developed a theory of algebra learning that emphasizes the importance of multiple strategies, comparison and explanation. Comparison of multiple strategies promotes learning in many domains, including mathematics and algebra in particular (Alfieri et al., 2013; Rittle-Johnson & Star, 2007, 2009; Rittle-Johnson, Star, & Durking, 2012; Gentner, Loewenstein, & Thompson, 2003). When paired with prompts to explain, comparing multiple strategies may be especially effective. Generating explanations improves learning across several topics and age groups (Aleven & Koedinger, 2002; Hodds et al., 2014; Rittle-Johnson, 2006).

Several short-term classroom studies have guided the development of our supplemental algebra curriculum comprised of worked example pairs and explanation prompts (Star, Pollack, et al., 2015). We designed worked example pairs and explanation prompts to support comparison of multiple strategies and whole-class mathematics discussions so that students had opportunities to generate explanations. Ongoing professional development was provided to support teachers' implementation of our curriculum, with a focus on facilitating mathematics discussions.

Method and Data Sources

Participants

The treatment condition consisted of nine Algebra I teachers and their students (n = 251) from three school districts in suburban Massachusetts who used our supplemental curriculum. One teacher taught 8^{th} grade and eight teachers taught 9^{th} grade. The business-as-usual control condition consisted of ten 9^{th} grade teachers and their students (n = 226) from one school district in suburban New Hampshire, who received the treatment the following year.

Design and Procedure

Teachers administered a pretest and posttest to their students at the beginning and end of the school year as well as for each of five target units. Treatment teachers used our materials during each of these five units, and data are reported for one of these units. In addition, a subset of lessons from both treatment and control classrooms were filmed and coded for key instructional practices described in more detail below.

Materials

Supplemental Curriculum. For the solving systems of equations unit, the treatment teachers were provided with nine supplemental lessons that were incorporated into the teachers' regular curricula. The materials were similar to those used in past research and showed the work of two hypothetical students who solved a math problem followed by prompts for explanation (Star, Pollack, et al., 2015, see Figure 1). Teachers also received feedback on the quality of their implementation of the previous unit and suggestions for improved use of our approach.

Assessment. We developed a unit assessment to measure students' knowledge of solving systems of equations. The 16-item assessment included three item types. Conceptual knowledge items (n = 5) targeted core concepts such as solutions to systems of equations (i.e., "Choose the definition that best describes the solution to a system of equations."). Procedural knowledge items (n = 6) required students to solve systems of equations (i.e., "If x + y = 12 and 2x + 5y = 36, what are the values of x and y?"). Procedural flexibility items (n = 6) primarily focused on identifying the most efficient strategy (e.g., "On a timed test, which would be the BEST way to solve the problem below?"). Cronbach's Alpha for the full assessment was .62 at pretest and .75 at posttest.

Video Coding Scheme. To investigate how use of our materials impacted teachers' instructional practices, videos were coded according to a detailed coding scheme adapted from a previous scheme on algebra instruction (Litke, 2015). Each video was broken into 7.5 minute segments to be coded because past work suggested that this was a short enough time to keep track of practices but long enough to see meaningful instructional interactions and have reliability (Litke, 2015). Details of the coding scheme are presented in Table 1.

Data Analysis

We compared student learning by testing differences in posttest scores between the treatment and control conditions while controlling for pretest scores. We used multilevel linear modeling to account for nesting of students in classrooms, as 31% of the total variance in posttest scores is due to classroom or teacher differences.

For video coding, calculated the mean rating for each instructional and classroom discussion quality coding category (e.g., the mean Interaction rating across video segments) and looked at the percentage of videos in which each comparison and written explanation activities occurred (e.g., what percentage of videos contained a comparison of multiple procedures) and. We did this for treatment teachers and separated segments within the same lesson where they used our materials from segments where they proceeded with their normal classroom instruction. Videos from the target unit were unavailable for control teachers, so we briefly describe their instructional practices using a small number of videos from other topics.

Results

Scores on the assessment improved from pretest to posttest for students in both conditions (see Figure 2). However, contrary to our expectations, students in the treatment condition had similar posttest scores as students in the control condition after controlling for pretest scores and classroom effects, $b = -.96$, $t(17.5) = -1.27$, $p = .22$. Pretest scores were a significant predictor of posttest scores, $b = .45$, $t(476.9) = 9.03$, $p < .001$.

Although students in the treatment condition did not outperform students in the control condition on the posttest, treatment teachers' instructional practices were desirable when using our materials. Of the 102 segments coded from 19 videos of treatment teachers' classroom lessons from this unit, 59% of the segments involved using our materials. Video coding results for treatment teachers are reported in Table 2. When teachers used our materials, they had average ratings of instructional content quality. Further, support for procedural flexibility was higher when teachers used our materials than when they did not use our materials. Importantly, teachers more often compared multiple procedures and better supported procedural flexibility when they used our materials than when they did not, which was the main focus of our materials. Ratings of teachers' classroom discussion quality were above-average, and students had many opportunities to provide written explanations. Use of our materials seemed to support the quality of treatment teachers' instructional content and especially their classroom discussions, relative to segments of their lessons when not using our materials.

We also describe coding of control teachers' videos when they were covering a different unit based on 44 segments coded from 9 videos, one per teacher. Control teachers' ratings for quality of instructional content and quality of classroom discussions were somewhat below average. They had average levels of making sense of procedures ($M = 2.20$) but low levels of supporting procedural flexibility ($M = 1.16$), and less than half of the videos contained a comparison of multiple procedures (44%). Control teachers' levels of teacher questioning, student responding, and interactions were similar and somewhat below average ($M = 2.00$ – 2.20). Control students were not asked to provide written explanations in any of the videos coded. Thus, control teachers' instructional practices from a different unit did not focus on supporting procedural flexibility or encouraging classroom discussions.

Significance

We found mixed evidence for the efficacy of our supplemental curriculum designed to promote comparison and explanation in algebra classrooms. Findings from coding of lesson videos suggest our materials impacted treatment teachers' instructional practices, at least compared to when they were not using our materials within the same lesson and compared to control teachers. Our materials supported the quality of teachers' instructional practices and classroom discussions as was intended. However, students' scores on both the pretest and posttest assessments were higher for students in the control classrooms than students in treatment classrooms. Unfortunately, the unexpected baseline differences in pretest scores between the two conditions make it difficult to interpret findings from our assessment data.

There is compelling evidence that comparison and discussion of multiple strategies should support student learning (e.g., Alfieri et al., 2013; Hodds et al., 2014; Woodward et al., 2012). However, supporting comparison and discussion in classrooms is difficult. Classroom discussions that allow students to generate explanations and build on one another's responses can improve learning (e.g., Lampert, 1990; Silver, Ghousseini, Gosen, Charalambous, & Strawhun, 2005), but teachers often struggle to ask more challenging, open-ended questions (Star, Newton, et al., 2015). In fact, teachers in the U.S. often have students explain simple aspects of comparison but explain the more challenging aspects themselves (Hiebert et al., 2003; Richland, Stigler, & Holyoak, 2012). We found that teachers only explicitly compared strategies when using our materials and that use of our materials increased teachers' levels of questioning. This indicates that teachers were using the comparisons of the worked example pairs as intended and confirmed past work that teachers are unlikely to explicitly compare strategies unless provided with supports to do so (Star, Pollack, et al., 2015). Further, students wrote explanations a majority of the time (72%) when using our materials but rarely did so when not using our materials. Thus, without explicit scaffolds to help students generate written explanations, such as our prompts, students were rarely given the opportunity to write down their own explanations. These findings suggest our supplemental curriculum was effective at supporting comparison and discussion in algebra classrooms, although this did not translate to higher posttest scores.

Our materials worked as intended to help teachers compare strategies, support procedural flexibility, and increase teacher questioning and student responding. However, mid-way through the school year, there was still room for growth in these levels of instructional practices. As we continue to process data from the remainder of the school year, we will examine how student learning and teachers' instructional practices improve as they continue to use our materials and receive professional development. It is likely that further supports may be needed to continue developing these practices in classrooms to improve students' learning.

References

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, *26*(2), 147–179. https://doi.org/10.1207/s15516709cog2602_1

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning Through Case Comparisons: A Meta-Analytic Review. *Educational Psychologist*, *48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). An Alternative Reconceptualization of Procedural and Conceptual Knowledge. *Journal for Research in Mathematics Education*, *38*(2), 115–131.

Blöte, A. W., Van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology*, *93*(3), 627–638.

Byrnes, J. P., & Wasik, B. A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology*, *27*(5), 777–786.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*, 393–405. https://doi.org/10.1037/0022-0663.95.2.393

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Human, P., Murray, H., … Wearne, D. (1996). Problem Solving as a Basis for Reform in Curriculum and Instruction: The Case of Mathematics. *Educational Researcher*, *25*(4), 12–21.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., … Kersting, N. (2003). Teaching mathematics in seven countries: Results from the TIMSS 1999 video study (NCES 2003-013). *US Department of Education. Washington, DC: National Center for Education Statistics*.

Hodds, M., Alcock, L., & Inglis, M. (2014). Self-Explanation Training Improves Proof Comprehension. *Journal for Research in Mathematics Education*, *45*(1), 62–101. https://doi.org/10.5951/jresematheduc.45.1.0062

Kieran, C. (1992). The learning and teaching of school algebra. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390–419). New York: Simon & Schuster.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, *27*, 29–63. https://doi.org/10.3102/00028312027001029

Litke, E. G. (2015). *The state of the gate: A description of instructional practice in algebra in five urban districts* (Doctoral dissertation). Harvard Graduate School of Education.

Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the Conceptual Structure of Mathematics. *Educational Psychologist*, *47*(3), 189–203. https://doi.org/10.1080/00461520.2012.667065

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*(1), 1–15. https://doi.org/10.1111/j.1467-8624.2006.00852.x

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*, 561-574. doi: 10.1037/0022-0663.99.3.561

Rittle-Johnson, B., & Star, J. R. (2009). Compared to what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*, 529-544. doi: 10.1037/a0014224

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2012). Developing procedural flexibility: Are novices prepared to learn from comparing procedures? *British Journal of Educational Psychology, 82*, 436-455. doi: 10.1111/j.2044-8279.2011.02037.x

Silver, E. A., Ghousseini, H., Gosen, D., Charalambous, C., & Strawhun, B. (2005). Moving from rhetoric to praxis: Issues faced by teachers in having students consider multiple solutions for problems in the mathematics classroom. *Journal of Mathematical Behavior*, *24*, 287–301. https://doi.org/10.1016/j.jmathb.2005.09.009

Star, J. R., Newton, K., Pollack, C., Kokka, K., Rittle-Johnson, B., & Durkin, K. (2015). Student, teacher, and instructional characteristics related to students' gains in flexibility. *Contemporary Educational Psychology*, *41*, 198–208. https://doi.org/10.1016/j.cedpsych.2015.03.001

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology*, *40*, 41–54. https://doi.org/10.1016/j.cedpsych.2014.05.005

Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, *18*, 565–579.

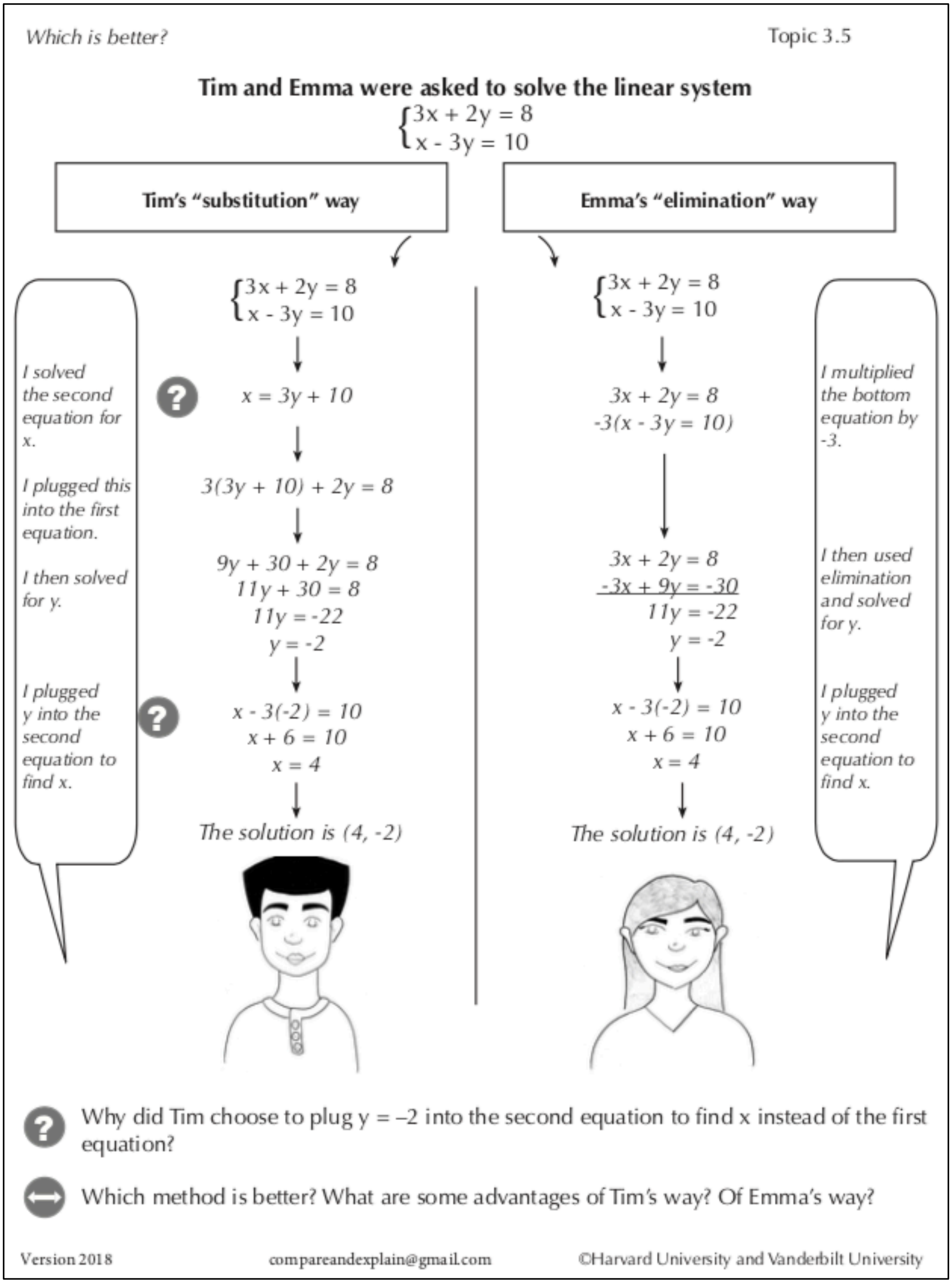
Woodward, J., Beckmann, S., Driscoll, M., Franke, M. L., Herzig, P., Jitendra, A. K., … Ogbuehi, P. (2012). *Improving mathematical problem solving in grades 4 through 8: A practice guide*. Washington, D. C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U. S. Department of Education.

Table 1. *Coding Scheme Categories*

Category	Description	Example of a 1 Rating	Example of a 4 Rating
Quality of Instructional Content			
Making Sense of Procedures	Teacher or students make meaning of individual steps of or a solution generated by a procedure, attend to the mathematical goals or properties underlying a procedure, or attend to why a procedure holds.	A procedure is presented like a recipe to be followed with no attention to meaning.	The focus of the segment is what it means for a system of equations to have infinite solutions.
Supporting Procedural Flexibility	The degree to which teachers present procedures to afford students the opportunity to notice multiple procedures to solve the same problem, attend to applicability of a procedure, attend to key conditions of steps within a procedure to understand its efficiency, or compare multiple procedures for their affordances and limitations.	A procedure is presented without any attention to its affordances/ limitations or alternative procedures.	Throughout the segment, the students compare when one procedure might be faster than another.
Quality of Classroom Discussions			
Teacher Questioning	The highest level question the teacher uses during the segment, and a student must respond in some way to the question in order for it to be considered. Question levels range from 1) "What" questions with simple one-word responses to 2) "How" questions about the steps of a procedure to 3) "Why" questions about understanding why an answer was correct or a procedure was a good choice to 4) "Open-ended" questions that allow students to share ideas, elaborate on others' ideas, and generalize.	A teacher asks, "What is the answer?"	A teacher asks, "Can you generate another problem where Riley's method could not be used?"
Student Responding	The highest level question or response a student generates during the segment. Response levels range from 1) answering with simple declarative information such as a numerical answer to 2) describing the steps of a procedure to 3) explaining their understanding such as justifying why a procedure is a good choice to 4) explaining a generalization or their understanding.	The student says, "The answer is x = 5."	The student states an explanation of why he or she does not agree with another student's choice of procedure.
Interactions	The highest level of student interaction defined as the opportunity to verbally share ideas regarding mathematical procedures and/or content, as elicited by the questioning.	The teacher is the only participant in the conversation.	Multiple students respond to the same question, and the teacher prompts them to build on each other's thinking.
Comparing Multiple Procedures	The teacher or students analyze multiple, different procedures for the same problem, noting the affordances or limitations of each. Rated as a Yes or No.	n/a	n/a
Use of Written Explanations	Students have an opportunity to produce written explanations. Rated as a Yes or No.	n/a	n/a

Table 2. *Mean Coding Ratings across Videos for Treatment Teachers for Systems of Equation Unit*

Code	Mean Rating (1 to 4 scale) With Materials	Mean Rating (1 to 4 scale) Without Materials
Quality of Instructional Content		
Making Sense of Procedures	2.20	2.43
Supporting Procedural Flexibility	2.57	1.33
Quality of Classroom Discussions		
Teacher Questioning	3.18	1.88
Student Responding	3.12	2.10
Interactions	2.97	1.88
Code	Percentage of Videos With Materials	Percentage of Videos Without Materials
Comparing Multiple Procedures	100%	37%
Use of Written Explanations	89%	5%



Which is better? Topic 3.5

Tim and Emma were asked to solve the linear system

$$\begin{cases} 3x + 2y = 8 \\ x - 3y = 10 \end{cases}$$

Tim's "substitution" way	**Emma's "elimination" way**
$\begin{cases} 3x + 2y = 8 \\ x - 3y = 10 \end{cases}$	$\begin{cases} 3x + 2y = 8 \\ x - 3y = 10 \end{cases}$
I solved the second equation for x. (?) $x = 3y + 10$	$3x + 2y = 8$ $-3(x - 3y = 10)$ *I multiplied the bottom equation by -3.*
I plugged this into the first equation. $3(3y + 10) + 2y = 8$	
I then solved for y. $9y + 30 + 2y = 8$ $11y + 30 = 8$ $11y = -22$ $y = -2$	$3x + 2y = 8$ $-3x + 9y = -30$ $11y = -22$ $y = -2$ *I then used elimination and solved for y.*
I plugged y into the second equation to find x. (?) $x - 3(-2) = 10$ $x + 6 = 10$ $x = 4$	$x - 3(-2) = 10$ $x + 6 = 10$ $x = 4$ *I plugged y into the second equation to find x.*
The solution is (4, -2)	*The solution is (4, -2)*

(?) Why did Tim choose to plug $y = -2$ into the second equation to find x instead of the first equation?

(↔) Which method is better? What are some advantages of Tim's way? Of Emma's way?

Version 2018 compareandexplain@gmail.com

Figure 1. *Sample Worked Example Pair*

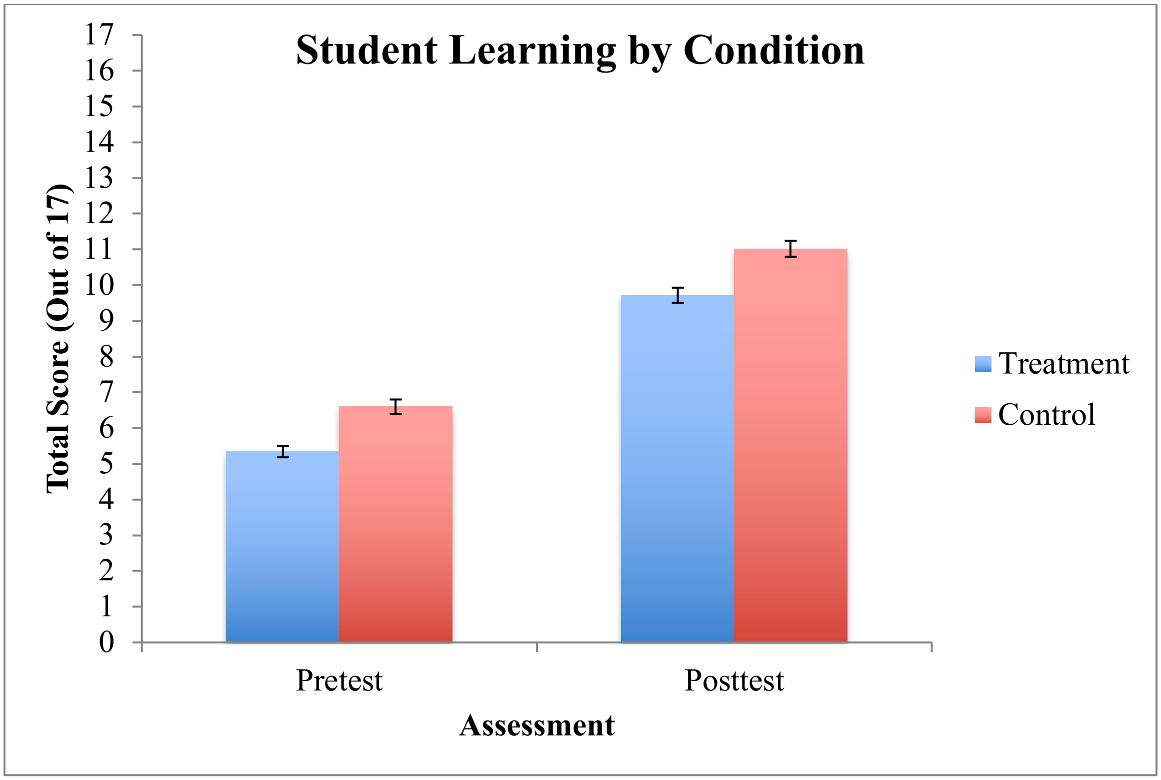


Figure 2. *Mean Total Scores at Pretest and Posttest by Condition*
Note. Error bars represent standard errors.