Journal of Educational Psychology
2016, Vol. 108, No. 2, 194–213

0022-0663/16/$12.00 http://dx.doi.org/10.1037/edu0000056

Teaching Mathematics by Comparison: Analog Visibility as a Double-Edged Sword

Kreshnik Nasi Begolli
University of California, Irvine

Lindsey Engle Richland
University of Chicago

Comparing multiple solutions to a single problem is an important mode for developing flexible mathematical thinking, yet instructionally leading this activity is challenging (Stein, Engle, Smith, & Hughes, 2008). We test 1 decision teachers must make after having students solve a problem: whether to only verbally discuss students' solutions or make them visible to others. Fifth grade students were presented with a videotaped mathematics lesson on ratio in which students described a misconception and 2 correct strategies. The original lesson was manipulated via video editing to create 3 versions with constant audio but in which the compared solutions were a) presented only orally, b) visible sequentially in the order they were described, or 3) all solutions were visible after being described throughout the discussion. Posttest and delayed posttest measures revealed the greatest gains when all solutions were visible throughout the discussion, particularly better than only oral presentation for conceptual knowledge. Sequentially showing students visual representations of solutions led to the lowest gains overall, and the highest rates of misconceptions. These results suggest that visual representations of analogs can support learning and schema formation, but they can also be hurtful—in our case if presented as visible in sequence.

Keywords: analogy, working memory, misconceptions, teaching strategies, video-lesson

Comparing different student solutions to a single instructional problem is a key recommended pedagogical tool in mathematics, leading to deep, generalizable learning (see Kilpatrick, Swafford, & Findell, 2001; National Mathematics Panel, 2008; Common Core Curriculum, 2013); however, the cognitive underpinnings of successfully completing this task are complex. In order to understand that $2 + 2 + 2$ conveys the same relationships as 2×3, for example, students must perform what has been theoretically described as structure-mapping: represent the multiple solutions as systems of mathematical relationships, align and map these systems to each other, and draw inferences based on the alignments (and misalignments) for successful schema formation (see Gentner, 1983; Gick & Holyoak, 1983). Structure-mapping is posited to underlie the processes of analogical reasoning where one *source* representation (e.g., $3 – 1 = 2$), is mapped to a *target* representation (e.g., $x – 1 = 2$), (Gentner, 1983).

Orchestrating classroom lessons in which learners successfully accomplish such structure mapping is not straightforward for many reasons. First, classroom discussions often involve comparisons between a misconception and a valid solution strategy, which may be particularly effortful in regards to structure mapping and schema formation, because misconceptions often derive from deeply or long-held beliefs that may be difficult to overcome (Vosniadou, 2013; Chi, 2013; Chinn & Brewer, 1993). Second, reasoners often fail to notice the relevance or importance of doing structure mapping unless given very clear and explicit support cues to do so (see Alfieri, Nokes-Malach, & Schunn, 2013; Gick & Holyoak, 1980, 1983; Gentner, Loewenstein, & Thompson, 2003; Ross, 1989; Schwartz & Bransford, 1998), Third, reasoners may intend to perform structure mapping but the process breaks down because their working memory or cognitive control processing resources are overwhelmed: (Cho, Holyoak, & Cannon, 2007; English & Halford, 1995; Morrison, Holyoak, & Troung, 2001; Richland, Morrison, & Holyoak, 2006; Waltz, Lau, Grewal, & Holyoak, 2000; Paas, Renkl, & Sweller, 2003). Working memory is required to represent the relationships operating within systems of objects as well as the higher order relationships between a familiar representation (*source analog*) and less familiar representation (*target analog*). In this case, to mentally consider the relationships between two solution strategies, one must hold in mind the steps to each solution strategy being compared, must reorganize and rerepresent these systems of relations so that their structures can align and map together, identify meaningful similarities and differences, and derive conceptual/schematic inferences from this structure-mapping exercise to better inform future problem solving (see Morrison et al., 2004; Morrison, Doumas, & Richland, 2011). Lastly, reasoners' prior knowledge plays an additional role

This article was published Online First July 20, 2015.

Kreshnik Nasi Begolli, School of Education, University of California, Irvine; Lindsey Engle Richland, Department of Comparative Human Development, University of Chicago.

This work would not have been possible without the support of the National Science Foundation CAREER grant NSF#0954222. We thank the participating schools, principals, teachers, and students. We also thank our lab manager Carey DeMichellis; research assistants Carmen Chan and James Gamboa; Alison O'Daniel for her help with Final Cut Pro; and Raminder Goraya for his programming efforts under high pressure.

Correspondence concerning this article should be addressed to Kreshnik Nasi Begolli, University of California, Irvine, School of Education, 3200 Education, Irvine, CA 92697-5500. E-mail: kbegolli@uci.edu

(Holyoak & Thagard, 1989; Holyoak, 2012). Those without adequate knowledge of the key relationships within the source and target representations are either unlikely to be able to notice structure mapping (Gentner & Rattermann, 1991; Goswami, 2001; Fyfe, Rittle-Johnson, & DeCaro, 2012), or this process will impose higher processing load than it would for those with more domain expertise (Novick & Holyoak, 1991).

These challenges mean that the instructional supports are very likely essential to whether students notice and successfully execute structure mapping between multiple solution strategies. The current study tests a classroom-relevant mode for providing such support—providing visual representations of the source and target analogs. The study manipulation assesses whether a) making source and target analogs visual (vs. oral) increases the likelihood that participants will notice and successfully benefit from structure mapping opportunities, and b) whether learning is enhanced if the visual representations of all compared solutions are visible simultaneously during structure mapping. The former should increase the salience of the relational structure of each representation, while the latter should reduce the working memory load and cognitive control resources necessary for participants to engage in structure mapping and inference processes.

Understanding the relationships between visual representations and learners' structure mapping provides insights into both a key pedagogical practice and improving theory on structure mapping and analogy more broadly. Teachers tend to find it difficult to lead students into making connections between problem solutions, and one productive way to support them is to provide guidelines for such discussions (e.g., see Stein, Engle, Smith, & Hughes, 2008).

Our study methodology is designed to lead to generalizable guidelines for the use of visual representations during classroom discussions comparing multiple solutions to a single problem. Observational data suggest that U.S. teachers do not regularly provide visual representations to support multiple compared solution strategies, and when they do, they are less likely than teachers in higher achieving countries to leave the multiple representations visible simultaneously (Richland, Zur, & Holyoak, 2007). The literature on the role of making representations visible suggests that presenting source and target analogs simultaneously versus sequentially leads to better learning (Gentner et al., 2003; Rittle-Johnson & Star, 2009; Richland & McDonough, 2010; Star & Rittle-Johnson, 2009), but these studies did not examine comparisons between an incorrect and a correct strategy. On the other hand, learning from incorrect and correct strategies was better than learning from correct strategies only (Durkin & Rittle-Johnson, 2012; Booth et al., 2013), but these studies have not investigated the role of visual supports. Thus this study may provide first evidence toward a guideline for teaching instructional comparisons with visual representations, particularly in the context of comparing a misconception and a correct student solution.

To maximize the relevance of our findings for teaching practices, we test alternative uses of visual representations within a mathematics lesson on proportional reasoning—a topic central to curriculum standards. Stimuli and data collection are conducted in everyday classrooms. The proportional reasoning lesson is situated in the context of a problem—asking students to find the best free-throw shooter in a basketball game. In this lesson, students are guided to perform structure mapping between three commonly used solution strategies: a) subtract between two units (e.g., subtract shots made from shots tried, which is incorrect and a common misconception), b) find the least common multiple between two ratios (e.g., proportionally equalize shots made to compare the shots tried), and c) divide two units to find a success rate (e.g., divide shots made by shots tried).

In addition, the work provides insights into theory on structure mapping and analogy. We examine a specific case of schema formation from structure mapping: identifying *misalignments* between two representations, in our case "subtraction" (a common misconception) and "proportions" (e.g., rate or ratio). To benefit from this structure-mapping exercise, students have to identify elements that are not aligned between the two relational structures. Namely, the difference between comparing a single unit (e.g., shots missed) and a relationship between two units (e.g., shots made and shots tried). Schema formation about proportional reasoning would derive from understanding the higher order differences between these two ways of attempting to solve the proportion problem. In contrast, structure-mapping failures may lead to the adoption of an inappropriate source (single unit comparison), or at best the target (relational comparison), but neither of which would be schema formation. In fact, either of these could hinder structure mapping, lead to misconceptions, and/or reduce transfer when solving later problems. We expect our findings to provide a more nuanced view on the possible implications of visual representations in terms of supporting or straining working memory resources necessary for successful structure mapping and its influence on students' mathematical knowledge.

We examine these research questions using an experiment that employs methods and measurements designed with the aim to optimize both ecological validity and experimental rigor. We utilize stimuli that approximate a true classroom experience—a single mathematics video-lesson recorded in a real classroom—then randomly assign students within each classroom to watch one of three versions of the lesson (see Figure 1). The recording is video-edited to support or strain working memory resources through variations in the visibility of representations. We use four carefully designed pre-post and delayed posttest measures to assess the impact of these manipulations on: 1) procedural understanding—students' ability to reproduce taught procedures; 2) procedural flexibility—participants' ability to understand multiple solutions and to deploy the optimal strategy; 3) conceptual understanding—understanding the concepts underpinning rate and ratio; and 4) use of misconceptions. These measures enable us to not only assess which use of visual representations is most effective for promoting learning, but they also let us better understand the processes by which children have been learning in each of the three conditions. Memory and retention of the instruction would be reflected in procedural understanding measures, while schema formation would be better reflected in the procedural flexibility and conceptual understanding measures. We theorize that working memory is the mechanism underpinning differences between these conditions on learning, since more working memory is required to hold visual representations in mind when reasoning about information that is not currently visible.

Thus, findings from this experiment will yield both theoretical insight into the role of visual representations for complex structure mapping, retention, and schema formation, and provides practice relevant implications for everyday mathematics teachers.

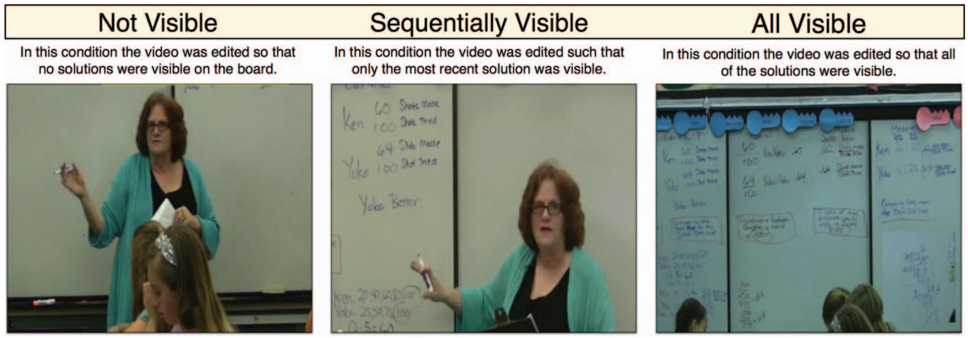


Figure 1. Still images illustrating the experimental conditions created through video editing. From left to right, the first picture shows only the teacher while obstructing the writing on the whiteboard (Not Visible condition), the middle picture shows only the most recent problem solution (Sequentially Visible condition), and the third picture shows the whole board (All Visible condition). See the online article for the color version of this figure.

Method

Participants

Eighty-eight participants were drawn from a suburban public school with a diverse population. Data from students who missed the intervention were omitted since their scores were not affected by our manipulation. Students who missed the pretest were also excluded from the analyses. They were excluded rather than having their data imputed (Peugh & Enders, 2004) due to concerns that solving pretest problems may have changed the learning context for those who took it due to a testing effect (Richland, Kornell, & Kao, 2009; Bjork, 1988; Carrier & Pashler, 1992; McDaniel, Roediger, & McDermott, 2007; Roediger & Karpicke, 2006a, 2006b; for a review see Richland, Bjork, Linn, 2007). The final analyses included 76 students (32 girls) who completed all three tests (pretest, immediate posttest, and retention test) with ages ranging between 11 and 12 years old.

Materials

Materials for the intervention consisted of a worksheet, a netbook, and a prerecorded video lesson embedded in an interactive computer program. Figure 2 provides a visual of the process for developing the lesson and administering it as stimulus to students in different schools.

Interactive Instructional Lesson

Proportional Reasoning. There is a large literature researching student thinking about ratio that has contributed to evidence that can predict students responses to proportion problems. (e.g.,

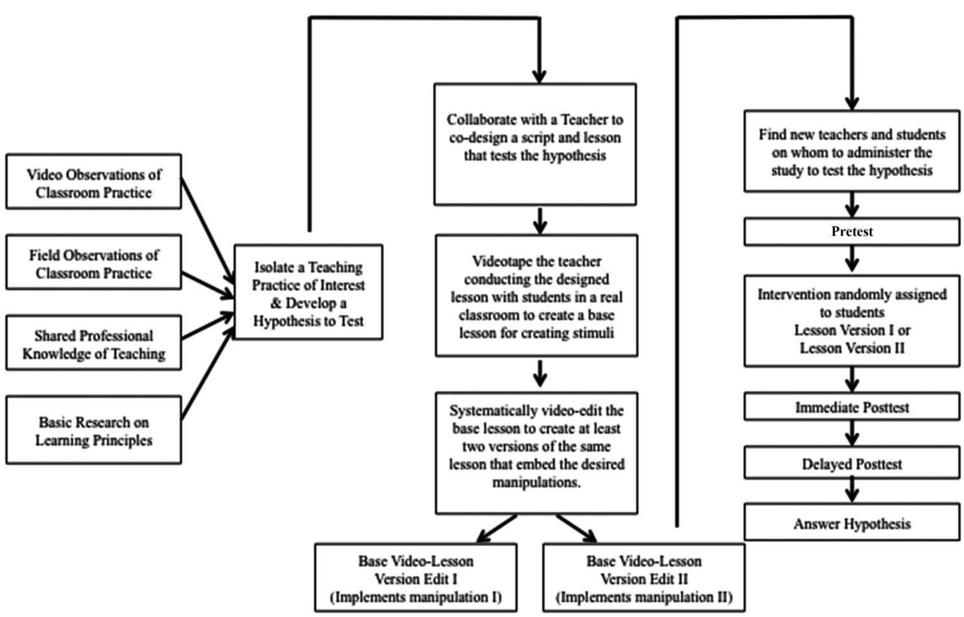


Figure 2. A process overview of using video to create experimental manipulations.

Hart, 1984; Hunting, 1983; Karplus, Pulos, & Stage, 1983; Kaput & West, 1994; Lamon, 1993a,1993b; Lo & Watanabe, 1997; Shimizu, 2003). Kilpatrick et al. (2001) identified proportional reasoning as requiring refined knowledge of mathematics and as the pinnacle of elementary arithmetic critical for algebraic and more sophisticated mathematics. Ratio was chosen for this study for two reasons: (a) it is part of the common core standards for sixth grade because it is essential for subsequent learning of algebra and (b) previous research has shown that ratio problems are cognitively taxing, leading to more diverse systematic student responses, useful for understanding mathematical thinking.

Lesson content and teacher collaboration. An approximately 40-min lesson was developed by the authors in collaboration with a nationally board certified public school teacher (see Appendix A for a sample of the transcript from the lesson). First, a lesson script was written based on a previously published lesson model (Shimizu, 2003) on which the teacher and the authors performed practice trials without students. During practice, the transcript was modified to feel more natural within the teacher's instructional style. The script provided specific details on how to present the main instructional problem; identify key student responses, present them on the board in a predetermined sequence, organize student responses on the board; the type of gestures to use, and so on. The teacher then taught the scripted lesson to her students in her regular classroom. Students were not given instructions and were expected to act spontaneously as they normally would during class hour. This teacher and students were not participants in our study. They only partook during the recording of the lesson which was used as stimulus for other students.

The lessons began with the teacher asking her students to solve the problem below.

Ken and Yoko were shooting free-throws in a basketball game. The results of their shooting are shown in the table below (Table 1). *Who is the better free-throw shooter?*

This was a novel problem and students were not given hints or instruction on how to solve it. The teacher's only instruction was to "solve any way you know how," and that "the class can learn from all the answers." If students objected because they did not know how to solve this, the teacher encouraged them to use any strategy they liked.

During the time when the students solved the problem, the teacher circled the room to identifying three students that used the three strategies listed in Table 2.

After a 5-min period, those three students were called to the board to share their solutions with the class. The sequence of strategies was presented based on this published lesson model (Shimizu, 2003). First, the incorrect solution C was verbalized by a student while the teacher wrote it on the board. Next, solutions B and then A were presented in the same manner with different students describing their strategies, followed by a short discussion on what the student was thinking when using the strategy. After all solutions were presented the teacher orchestrated a discussion comparing the different solution methods to achieve specific goals. The teacher's goals were: (a) to challenge students' common misconception (strategy C—subtraction) by asking students whether strategy C is reasonable if the numbers changed (Ken makes 0 out of 4 free throws and Yoko makes 5 out of 10), (b) introduce the concept of proportional reasoning (strategy B) by leading students to notice that proportions can be compared by making one number (i.e., the denominator) constant in each ratio and then comparing the other number (i.e., the numerator) to determine which ratio is larger, (c) to challenge students to notice that the least common multiple strategy can become more difficult for larger and prime numbers, (d) to notice that using division (solution A) is the most efficient strategy, since it does not change much in difficulty, regardless if the numbers increase. These points were orchestrated by the teacher through predesigned comparisons that led her to introduce the concept of ratio, while the class responded spontaneously to her prompts.

Table 1
Problem Solved and Discussed During Videotaped Lesson

	Shots made	Shots tried
Ken	12	20
Yoko	16	25

The crux of our manipulation came from applying video-editing techniques to the recording to create three different versions of the same lesson. FINAL CUT PRO's (FCP) 7.0.3 academic version's various editing features were used such as zooming, cropping, or different camera perspectives of the screen canvas to either: (a) hide the board to create a version of the lesson for the Not Visible condition; (b) show only the section of the board most recently discussed, but hide other areas of the board to create a version of the lesson for the Sequentially Visible condition; or (c) show the whole board throughout the version of the lesson in the All Visible condition. Thus, the same content was verbalized in all three lesson versions, but with systematic differences in visual cues.

Each version of the lesson was strategically divided into nine clips with an approximate range from 1 min to 8 min. The endpoints of each clip were chosen based on when the teacher asked questions to the class. Each version of the video-lesson was made interactive by embedding clips of the video in a computer program written specifically for this study. At the end of each clip, the program prompted students with questions that were asked by the teacher in the videotaped classroom. Students in all conditions either wrote their answers on a packet provided by the experimenters, or selected multiple choice questions that the computer program collected as assessment data. This methodological approach of stimuli creation, provided a rigorous level of experimental control of a highly dynamic context—an everyday classroom. Further, it allowed for randomization within each classroom.

Assessment

The assessment was designed to assess schema formation and generalization. Mathematically, the assessment included three constructs, procedural knowledge, procedural flexibility, and conceptual knowledge (Rittle-Johnson & Schneider, in press). These constructs were conceptually derived from Rittle-Johnson and Star (2007, 2009), and adapted to the core concepts and procedures underlying ratio problems. Items used for assessing each knowledge type are included in Appendix B. The items on the pretest and posttests were identical, but the pretest contained five additional procedural knowledge problems used to assess students' prerequisite knowledge of basic procedures (e.g., division by decimals, finding the least common denominator). Detailed scoring on all of

Table 2
Three Student Solutions Compared During the Videotaped Instructional Analogy Lesson

A	Student finds the number of goals made if each player shoots only 1 free throw. Ken: 12 goals ÷ 20 shots = .6, and Yoko: 16 goals ÷ 25 shots = .64. *Answer: Yoko, because she gets more goals for the same number of free throws (.64 > .60).*	Most efficient generalizable strategy
B	Student compares the number of goals if each player shoots the same number of free throws. Using 100 as the last common multiple, we get Ken: 60/100 and Yoko: 64/100. *Answer*: *Yoko, because she would get more goals if they each shot 100 times (64/100 > 60/100).*	Finding least common multiple: Drawback, difficult when larger numbers
C	Student compares the players by finding the difference between the number of free throw shots and the number of goals. Ken: 20 shots − 12 goals = 8 misses, and Yoko: 25 shots − 16 goals = 9 misses. *Answer: Ken, because he missed fewer times than Yoko (8 < 9).*	Misconception (incorrect): subtract values and compare differences without considering the ratio.

the items can be found in Appendix B. Scores for each construct were averaged to yield an overall mean for that particular construct. Student produced strategies were coded for use of: (a) division, (b) least common multiple, (c) subtraction, (d) ratio, (e) other valid strategy (e.g., cross multiplication), (f) other invalid strategy (e.g., addition). For the analyses, these codes were collapsed for each problem into correct or incorrect. Open-ended questions (e.g., what is the definition of ratio?) were given a binary score of either correct or incorrect. Interrater agreement for three raters was calculated on 20% of each test phase on whether the strategy was correct or incorrect ranging from 95–99% (0.88–0.97 Kappa). Strategy specific coding was less reliable ranging from 86%–92% (0.79–0.85 Kappa).

Procedural Knowledge. Seven problems measured baseline ability and growth in procedural understanding. The procedural knowledge construct was designed to test: (a) students' knowledge for producing and evaluating solutions of: familiar problems (i.e., similar in appearance to the problem used in the video lesson; $n = 2$ produce; $n = 3$ recognition), (b) transfer problems (i.e., students' competence to extend these solution strategies to problems with novel appearance, but similar context; $n = 2$). Students received one point if they produced the correct solution method and another point if they produced a correct solution method and made an inference to reach a correct answer. Students received one point if they recognized the correct strategy on multiple-choice questions. One procedural transfer problem showed no sensitivity to the intervention, so it was dropped from the analyses (average change of 5% from pretest to posttest; see Problem 4 in Appendix B for details). Cronbach's alpha on the remaining items was .88 at posttest, .91 at delayed posttest, and .86 at pretest, which are above the suggested values of .5 or .6 (Nunnally, 1967).

Procedural flexibility. The procedural flexibility construct measured: (a) students' adaptive production of solution methods ($n = 3$), (b) their ability to identify the most efficient strategy ($n = 1$), and (c) students' ability to identify a novel solution method which was related to a taught strategy ($n = 1$). For (a) students were presented with one problem containing three items. The first item asked students to produce two strategies (and correct answers) for the same problem. The second item asked students to evaluate which of the two strategies was most effective. The third item asked students to select one out of four reasons for their choice on Item 2. Students could receive two points for the first item and one point on the last two items. For (b) students were presented a multiple-choice problem that required them to identify the optimal strategy from two valid and two invalid strategies. For (c) students were presented with a multiple choice problem that probed students' competence to identify the correctness of a related but novel method of solving a problem (i.e., finding the lowest common multiple for the numerator instead of the denominator). Both (b) and (c) were scored for accuracy. Cronbach's alpha on the flexibility construct was .68 at posttest, .67 at delayed posttest, and .57 at pretest.

Conceptual knowledge. The conceptual knowledge construct consisted of seven items that were designed to probe students' explicit and implicit knowledge of ratio. Students' explicit knowledge was measured by asking them to write a definition for ratio, which was scored for accuracy. The other six items measured students' implicit understanding and transfer to new contexts. One problem probed whether students could conceptually examine two sets of non-numerical quantities (i.e., pictures of lemon juice and water), adapt their just learned solution methods to this novel context, and compare the sets to decide which ratio was greater (i.e., which lemonade was more lemony?). On this problem, students were scored on whether they could produce the correct setup given objective quantities and choose the correct set (when using the correct setup). The remaining problems were scored for accuracy. One multiple-choice item that was part of a procedural knowledge problem probed whether students could conceptually evaluate the multiplicative properties of a solution procedure they had produced (see Problem 1 in Appendix B under the Procedural and Conceptual Knowledge sections). Three conceptual questions, two fill-in-the-blank, and one multiple choice, probed students' understanding of units in correspondence to ratio and rate numerical quantities. One of these unit questions was dropped due to floor effects (only 6 out of 97 students responded correctly; see Problem 5, Part 3 in Appendix B for details). Lastly, one problem asked students to recognize the correct solution and setup for a novel problem type and context. However, this problem suffered from ceiling effects and was not sensitive to intervention (averages ranged between 81%–89%; average change −3% from pretest to posttest; see Problem 8 in Appendix B for details). Thus, it was omitted from the analyses. Cronbach's alpha on the remaining items was .66 at posttest, .64 at delayed posttest, and .42 on pretest. Reliability at pretest was lower due to floor effects.

Common misconception. Misconceptions are mistakes that students make based on inferences from prior knowledge, which obstruct learning (Smith, diSessa, Roschelle, 1993). Based on a published lesson (Shimizu, 2003), pilot data, and pretest data, a solution involving subtraction was expected to be the most common misconception participants would bring to the study. A subcoding assessed how frequently students used the subtraction method. The common misconception measure examined students' use of subtraction on near transfer procedural problems that looked like the instructed problem in the video lesson.

Design and Procedure

Students within four classrooms, not in the videotaped classroom, were randomly assigned to three experimental conditions: Not Visible ($n = 26$), Sequentially Visible ($n = 26$), or All Visible ($n = 24$). All students were administered a pretest, 1 week later completed the video-lesson intervention and an immediate posttest, and o1 week later completed a delayed posttest. Students underwent the intervention before being introduced to rate and ratio in their regular curriculum.

Results

Baseline Data

One-way ANOVAs were conducted first to establish that the randomization was successful and there were no differences between conditions on each of the above described constructs. At pretest, there were no differences between conditions on any of the outcome constructs: procedural, procedural flexibility, conceptual knowledge, and common misconception with all p values above .05, Fs (2, 80) = 0.69, 0.53, 0.71, and 0.29, respectively. At pretest, students used mostly invalid strategies when solving ratio problems, and left a significant proportion of the problems blank (see Table 3). The average scores at each test point by condition are summarized in Table 4.

Condition Effects

Analysis plan overview. We next sought to examine the effects of condition on each of the dependent variable constructs measured. There were three primary constructs: procedural knowledge, procedural flexibility, and conceptual understanding, and one additional measure to gather deeper information on the impact of the manipulations on inappropriate retention—use of the misconception.

We conducted separate ANCOVAs for each outcome measure with both posttests as a within-subjects factor (immediate test and delayed test) and condition as a between-subjects factor. Students' pretest accuracy and their classroom (i.e., teacher) were included as covariates. In the model, the pretest measure matched the posttest measure, such that procedural knowledge pretest served as a covariate for the procedural knowledge posttests, and so on. Levene's test of variance homogeneity was used to ensure that all measures were appropriate for use of the ANCOVA statistic. These analyses yielded no significant differences in variance between groups on all measures (F values range $.078 < F < 1.764$ and p value range $.173 < p < .925$), apart from the score for how often students used the misconception. The measure of misconception use was therefore analyzed using Mann–Whitney U comparisons of pretest to posttest gain scores across conditions, with a Bonferroni correction for multiple comparisons.

When a main effect of condition was present on an ANCOVA analysis, least significant difference tests were used to determine whether there were differential effects of condition on posttest performance. Student performance was not expected to change between posttests because students continued to learn about ratio-related concepts after the intervention and our within-subject test for time confirmed this prediction.

Main effects of condition. The results of each ANCOVA are summarized in Table 5. For each outcome there was a main effect of condition with moderate to high effect sizes ($.11 < \eta^2 < .15$) and sufficient power ($.77 < (1-\beta) < .90$). Pretest was a significant predictor for each construct, though misconception use was not independently predictive. There were no expectations that time of test or classroom teacher would interact with condition and our tests support this. Pairwise comparisons between conditions on each construct are reported below (see Table 6 and Figure 3).

Procedural knowledge. Students in the All Visible condition outperformed students in the Sequentially Visible condition in procedural knowledge. An unexpected finding was that students in the Not Visible condition also outperformed students in Sequentially Visible condition (see Table 6). Not seeing the board did not affect students' procedural knowledge compared to students who saw all solutions on the board simultaneously.

Table 3
Solution Strategies Produced for Ratio Problems by Condition

	Blank	Division	Least common multiple	Ratio setup	Subtraction	Other valid	Other invalid
Pretest							
All visible	25%	4%	18%	9%	20%	7%	18%
Seq. visible	24%	8%	7%	4%	21%	4%	30%
Not visible	17%	10%	10%	8%	22%	2%	30%
Posttest							
All visible	14%	48%	22%	1%	13%	3%	4%
Seq. visible	12%	19%	15%	1%	29%	1%	17%
Not visible	5%	39%	21%	1%	21%	1%	14%
Delayed							
All visible	16%	39%	22%	1%	10%	3%	6%
Seq. visible	15%	25%	12%	0%	32%	5%	11%
Not visible	12%	30%	19%	4%	19%	4%	12%

Table 4
Student Scores, by Condition

	Pretest		Posttest		Delayed	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Procedural						
All visible	33%	0.29	59%	0.37	67%	0.35
Sequentially visible	25%	0.28	38%	0.35	38%	0.35
Not visible	27%	0.32	52%	0.34	56%	0.35
Flexibility						
All visible	11%	0.14	37%	0.29	42%	0.26
Sequentially visible	14%	0.22	17%	0.24	22%	0.29
Not visible	10%	0.14	23%	0.25	30%	0.31
Conceptual						
All visible	19%	0.19	44%	0.29	47%	0.26
Sequentially visible	13%	0.17	30%	0.30	27%	0.28
Not visible	17%	0.22	31%	0.25	35%	0.23
Common misconception						
All visible	20%	0.21	13%	0.20	10%	0.15
Sequentially visible	21%	0.21	29%	0.31	32%	0.27
Not visible	22%	0.22	21%	0.24	19%	0.24

Procedural flexibility. Students in the All Visible condition outperformed students in the Sequentially Visible condition, but there were no differences between any other groups (see Table 6).

Conceptual knowledge. Pairwise comparisons reveal that students in the All Visible condition scored significantly higher than students in the Not Visible condition and students in the Sequentially Visible condition (see Table 6). There were no differences between students in the Sequentially Visible and Not Visible condition on conceptual knowledge (see Table 6).

These results could have been driven by the types of solution strategies that students used (Rittle-Johnson & Star, 2007, 2009), particularly subtraction (a common misconception).

Common misconception. The general pattern for students' use of the common misconception displays a reverse pattern compared to the procedural knowledge performance, providing insight into why the Sequentially Visible condition led to low accuracy rates. Mann–Whitney U pairwise comparisons with a Bonferroni adjusted alpha level of p = .0167 per test (.05/3) show that students in the Sequentially Visible condition used the common misconception significantly more from pretest to delayed test compared to students in the All Visible condition, but no other comparisons were significant (see Table 7).

Discussion

Overall, this study supported our hypothesis that the presence of visual representations during a discussion comparing multiple solutions to a problem can serve as a double-edged sword. The presence and timing of visual representations impacted children's learning from a mathematical classroom lesson on ratio when comparing a misconception to two correct strategies in both positive and negative ways. Having all visual representations available simultaneously led to the highest rates of learning, while having them presented sequentially led to the highest rates of misconceptions.

Specifically, the ability to see all compared representations simultaneously throughout the discussion increased the likelihood of schema formation and optimized learning when compared with seeing compared representations only sequentially. This was evidenced by greater ability to: a) use taught procedures, b) understand multiple accurate solution strategies and select the most efficient strategy, c) explain and use the concepts underlying taught mathematics, and d) minimize use of a misconception.

Strikingly, presenting mathematical solutions sequentially led to the lowest performance on these positive measures of learning, and the highest rates of misconceptions at posttests. This condition led to even lower learning rates overall than having no visual representations present during any of the comparison episodes, though these differences were not present on all measures. The details of how these conditions differed are informative to building theory regarding the role of visual representations in comparisons and schema formation.

Having the solution strategies presented only verbally (Not Visible condition) led to performance rates that fell in between the two visual representation conditions. Not Visible presentation did lead to some retention of taught procedures and some schema formation, but not as universally as in the All Visible condition. At the same time, these participants (Not Visible condition) were less likely than in the Sequential condition to produce the misconception, suggesting that they did not retain the instructed representations as well or uncritically as in that condition. It may be that the Not Visible condition was most effortful for students and thus some students were less successful than in the All Visible condition, but for those students who were able to perform that effort, their learning was strong.

Drawing on theory on the cognitive underpinnings of structure mapping, we interpret the differences between these conditions based on their likely load on students' executive function resources. Structure mapping is well established to require both the ability to hold representations in mind and manipulate the relationships to identify and map structural alignments or misalignments (e.g., Waltz et al., 2000; Morrison et al., 2004), as well as to effortfully inhibit attention to invalid relationships (e.g., Cho et al., 2007; Richland & Burchinal, 2013). We suggest that having all

Table 5
Analyses of Covariance Results on Learning Outcomes

Factor	Procedural knowledge				Procedural flexibility				Conceptual knowledge			
	F	MSE	p	η^2	F	MSE	p	η^2	F	MSE	p	η^2
Condition	4.74	.67	.012	.12	5.68	.660	.005	.14	4.41	.37	.016	.11
Pretest	37.79	5.32	.000	.35	4.49	.522	.038	.06	29.33	2.44	.000	.30
Teacher	2.68	.38	.106	.04	0.97	.113	.327	.01	0.66	.05	.420	.01

* Condition degrees of freedom are (2, 66); all others are (1, 70).

Table 6
Pairwise Comparisons, by Condition With Pretest and Teacher as Covariates for Both Posttests

Knowledge type	AV vs. SV	AV vs. NV	SV vs. NV
Procedural knowledge	**.007**	.778	**.013**
Procedural flexibility	**.001**	.129	.071
Conceptual knowledge	**.005**	**.041**	.407

Note. The numbers above reflect *p*-values. The numbers in bold are *p*-values $< .05$.

Table 7
Summary of Mann-Whitney U Pairwise Comparisons of the Common Misconception

	Immediate posttest		Delayed posttest	
	Mann Whitney U	*p*-value	Mann Whitney U	*p*-value
AV vs. SV	208	0.034	178.5	0.007
AV vs. NV	277.5	0.49	270.5	0.41
SV vs. NV	263	0.159	266	0.179

Note. Bonferroni adjustment renders alpha levels at $p = .0167$ (.05/3).

visual representations available during structure mapping reduced the working memory load required for participants to hold the representations active in mind, so they could use those resources more directly for structure mapping.

In contrast, we suggest that having the representations presented sequentially may have imposed the highest burden on the executive function system, requiring students to effortfully inhibit attention to the misconception representation presented first. This representation was likely salient for its visual cues as well as for its coherence with prior knowledge (hence being a common misconception). So, suppressing the impulse to retain and use this representation as it was and rather to rerepresent this information through structure mapping may have been particularly effortful and thus successful less of the time. Performing analogical reasoning would have required executive function resources to revisit the misconception in light of subsequent strategies to discard its validity. However, for the Sequentially Visible group the misconception was no longer visible throughout the comparison; thus, making it more difficult to identify misalignments between the appropriate strategy and the misconception.

As dual coding theory would suggest, reinforcing exemplars through visual and auditory presentations leads to greater retention (Clark & Paivio, 1991). Higher retention for the details of presented representations might explain why participants in the sequential condition were most likely to retain and produce the misconception at posttest, rather than showing evidence of schema formation—which would have been expected if the students performed structure mapping. There is significant literature suggesting that people tend to use data in the world to confirm their biases, potentially leading toward difficulties in drawing appropriate inferences from analogies (Brown, 2014; Zook, 1991). Thus in this case this confirmation bias seems to have led participants to retain the misconception as presented.

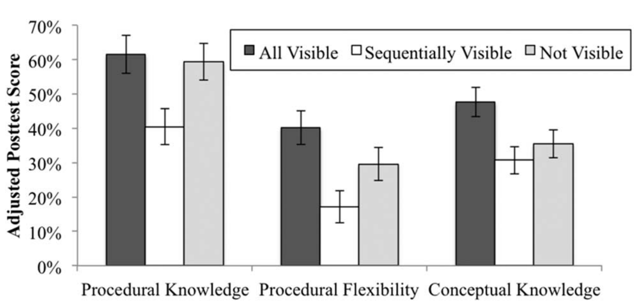


Figure 3. Estimated marginal means across both posttests on procedural knowledge, procedural flexibility, and conceptual knowledge by condition. Error bars are standard errors.

In sum, these results provide insight into the role of visual representations in schema formation. Presence of visual representations can aid structure-mapping and schema formation when representations of all compared solutions are visible, in particular to improve conceptual understanding. However, having visual representations presented only sequentially can actually hinder structure mapping, leading to retention of the details of the representations rather than the overarching schema. This is particularly evident in the situation tested here, in which one representation being compared is a common misconception. Presenting analogs sequentially increased usage of the misconception on posttest when compared to having the analogs presented simultaneously, which suggests that the visibility of the analogs may play an important role in either supporting or derailing structure mapping.

Implications for Theory and Practice

The findings from this study have the potential to inform U.S. teaching practices as well as to contribute to several areas of cognitive scientific literatures. From a theoretical standpoint, these findings extend previous laboratory-based results on analogical learning to classroom contexts, using a video-based methodology with high ecological validity.

In addition, the work extends studies of visual representations to examine the role of visual representations on schema formation when relational analogs include a misconception. Misconceptions are mostly unexplored in prominent structure-mapping models (Gentner, 1983; Gentner & Forbus, 2011). Conceptual change literature (Vosniadou, 2013; Chi, 2013; Carey & Spelke, 1994) has investigated how people overcome misconceptions in the context of science education (Chinn & Brewer, 1993; Brown & Clement, 1989; Brown, 2014), and recently in mathematics (Vamvakoussi & Vosniadou, 2012), but the influence of misconceptions on structure-mapping models remains to be fully defined. So, this study has potential to contribute to both the conceptual change and analogy literatures.

These results are also informative in moving toward recommendations for teachers regarding optimal use of visual representations during instructional comparisons—particularly for leading discussions about multiple ways of solving single problems. From an instructional perspective, showing visual representations and making mathematical comparisons is common to everyday mathematics instruction (Richland et al., 2007). Thus shifting to leave all source and target representations visible throughout a full mathematical discussion, rather than only while they are first being presented, requires only a reorganization of existing routines rather than a large time investment and modification of current practice. Thus this intervention is feasible for integration into current teaching practices.

Additionally, the use of more ecologically valid stimuli to test teaching practices through a videotaped teacher guided lesson, instead of static written learning materials, ideally allows for greater generalizability of our findings. Though, we warn against interpreting these results to indicate that making analogs visible simultaneously will always lead to successful structure mapping and mathematical schema formation. Only that if the analogs being compared are informative and the learner notices their relationship does making them visible simultaneously aid in abstraction.

A primary constraint to implementation of making representations visible throughout lessons is space. When codesigning this lesson with teachers we faced the challenge that teachers often use their presentation space (e.g., white boards) for many purposes including daily schedules and reminders, which may reduce the amount of space available to leave multiple representations visible. This challenge is compounded by the trend to reduce presentation space through the use of such technologies as electronic whiteboards, such as *innovative white board* technologies (IWB; De Vita, Verschaffel, Elen, 2014). These innovations enable teachers to control the board from their computer in a dynamic fashion, allowing for advanced preparation or careful design of visual representations, which can be a great strength. However there is also typically less room to make multiple representations visible, because these boards are about a third of the size of typical classroom chalk or white boards. These data suggest that IWBs (e.g., Smart boards) have the potential to be highly effective at instantiating single visual representations at a time, much as in our sequentially visible condition, which led to the lowest learning gains and greatest rate of misconceptions. Thus, our data imply that teachers could enhance learning by invoking creativity in using these technological options to make a record of multiple visual representations.

In summary, these findings suggest that instructional recommendations should emphasize the utility of making compared representations visible simultaneously, but more broadly to highlight the importance of supporting learners in aligning, mapping, and drawing inferences about the similarities and differences across representations such as multiple solution strategies for a problem. Teachers should also be made aware of the challenges inherent in making such comparisons when one of the representations is a misconception. In such cases, students may need additional support to control their attentional responses to the misconception in order to engage in more productive knowledge rerepresentation and new schema formation. In the context of instructional analogies, it is important to consider that visual representations should highlight relationships between representations, not just increase salience, memorability, and clarity of one representation. The latter has the potential to support deeper encoding of a misconception, rather than desired schema formation that leads to generalizable learning.

Limitations and Future Directions

The current study provides important findings on the role of visual representations for challenging a common misconception through structure-mapping in the mathematical area of proportional reasoning. While this is an area that is critical for students' future attainment of algebra (Kilpatrick et al., 2001), a broader variety of mathematical domains need to be tested to examine the universality of our results before making a clear guideline for teachers.

A strength of our study is that the instructional stimuli derive from videodata of a real classroom lesson, leading to a simulation of an everyday classroom learning experience, with the aim to increase the study's generalizability to teaching practices. While video lessons are an increasing trend with the heightened use of methodologies such as "flipped classrooms" (Jinlei, Ying, & Baohui, 2012) in higher education, elementary students generally interact with live teachers, instead of recordings of a teacher. Despite this, video can convey emotion, body language, and other nonverbal cues, thus offering a more realistic medium than text-based or computerized materials. Further, teacher actions within a video lesson are more translatable to a true lesson.

Thus, this technology has high potential for maximizing internal and external validity for testing findings evidenced in laboratory contexts and translating them to teaching practices as well as isolating the efficiency of instructional methods that teachers routinely use in their classrooms. At the same time, there are limits to the simulation, so a future direction for this work would be to extend the methodology into testing teacher-delivered material. Additional future directions include using the video methodology to test the efficacy of additional aspects of the instructional routines to provide additional explicit guidelines, including use of teacher gestures or order of presenting contrasting representations.

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48,* 87–113. http://dx.doi.org/10.1080/00461520.2013.775712

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory II* (pp. 396–401). London, UK: Wiley.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction, 25,* 24–34.

Brown, D. E. (2014). Students' conceptions as dynamically emergent structures. *Science & Education, 23,* 1463–1483. http://dx.doi.org/10.1007/s11191-013-9655-9

Brown, D. E., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: Abstract transfer versus explanatory model construction. *Instructional science, 18,* 237–261.

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In. L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 169–200). Cambridge, UK: Cambridge University Press.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633–642. http://dx.doi.org/10.3758/BF03202713

Chi, M. T. H. (2013). Two kinds and four sub-types of misconceived knowledge, ways to change it, and the learning outcomes. In S. Vosniadou (Ed.), *The international handbook of conceptual change* (2nd ed., pp. 49–70). New York, NY: Routledge. http://dx.doi.org/10.4324/9780203154472.ch3

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1–49. http://dx.doi.org/10.3102/00346543063001001

Cho, S., Holyoak, K. J., & Cannon, T. D. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Memory & Cognition, 35,* 1445–1455. http://dx.doi.org/10.3758/BF03193614

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3,* 149–210. http://dx.doi.org/10.1007/BF01320076

Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

De Vita, M., Verschaffel, L., & Elen, J. (2014). Interactive whiteboards in mathematics teaching: A literature review. *Education Research International 2014, Article ID 401315*. http://dx.doi.org/10.1155/2014/401315

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22,* 206–214. http://dx.doi.org/10.1016/j.learninstruc.2011.11.001

English, L. D., & Halford, G. S. (1995). *Mathematics education: Models and processes*. Mahwah, NJ: Erlbaum.

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology, 104,* 1094–1108. http://dx.doi.org/10.1037/a0028389

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170. http://dx.doi.org/10.1207/s15516709cog0702_3

Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science, 2,* 266–276. http://dx.doi.org/10.1002/wcs.105

Gentner, D., & Rattermann, M. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). Cambridge, UK: Cambridge University Press.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95,* 393–405.

Gick, M. L., & Holyoak, K. L. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306–355. http://dx.doi.org/10.1016/0010-0285(80)90013-4

Gick, M. L., & Holyoak, K. L. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1–38. http://dx.doi.org/10.1016/0010-0285(83)90002-6

Goswami, U. (2001). Analogical reasoning in children. In D. Gentner, K. J. Holyoak, and B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 437–470). Cambridge: MIT Press.

Hart, K. M. (1984). *Ratio: Children's strategies and errors*. Windsor, UK: NFER-Nelson.

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 234–259). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199734689.001.0001

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295–355.

Hunting, R. P. (1983). Emerging methodologies for understanding internal processes governing children's mathematical behaviors. *Australian Journal of Education, 27,* 45–67.

Jinlei, Z., Ying, W., & Baohui, Z. (2012). Introducing a new teaching model: Flipped classroom. *Journal of Distance Education, 4,* 46–51.

Kaput, J., & West, M. (1994). Missing value proportional reasoning problems: Factors affecting informal reasoning patterns. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 61–85). Albany, NY: SUNY Press.

Karplus, R., Pulos, S., & Stage, E. (1983). Proportional reasoning of early adolescents. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 45–90). Orlando, FL: Academic Press.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics* (Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education, National Research Council). Washington, DC: National Academy Press. Retrieved from http://www.nap.edu/catalog/9822.html

Lamon, S. (1993a). Ratio and proportion. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 131–156). Hillsdale, NJ: Erlbaum.

Lamon, S. (1993b). Ratio and proportion: Connecting content and children's thinking. *Journal for Research in Mathematics Education, 24,* 41–61. http://dx.doi.org/10.2307/749385

Lo, J.-J., & Watanabe, T. (1997). Developing ratio and proportion schemes: A story of a fifth grader. *Journal for Research in Mathematics Education, 28,* 216–236. http://dx.doi.org/10.2307/749762

McDaniel, M. A., Roediger, H. L., III, & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14,* 200–206. http://dx.doi.org/10.3758/BF03194052

Morrison, R. G., Doumas, L. A. A., & Richland, L. E. (2011). A computational account of children's analogical reasoning: Balancing inhibitory control in working memory and relational representation. *Developmental Science, 14,* 516–529. http://dx.doi.org/10.1111/j.1467-7687.2010.00999.x

Morrison, R. G., Holyoak, K. J., & Truong, B. (2001, August). *Working memory modularity in analogical reasoning*. Presentation given at the 23rd Annual Conference of the Cognitive Science Society, Edinburgh, UK.

Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience, 16,* 260–271. http://dx.doi.org/10.1162/089892904322984553

National Mathematics Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U. S. Department of Education.

Nunnally, J. C., Bernstein, I. H., & Berge, J. M. T. (1967). *Psychometric theory* (Vol. 226). New York: McGraw-Hill.

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 398–415.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38,* 1–4. http://dx.doi.org/10.1207/S15326985EP3801_1

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74,* 525–556. http://dx.doi.org/10.3102/00346543074004525

Richland, L. E., Bjork, R. A., & Linn, M. C. (2007). Cognition and instruction: Bridging laboratory and classroom settings. In F. Durso, R. Nickerson, S. Dumais, & T. Perfect (Eds.), *Handbook of applied cognition* (pp. 555–583). Chichester, UK: Wiley. http://dx.doi.org/10.1002/9780470713181.ch21

Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science, 24,* 87–92. http://dx.doi.org/10.1177/0956797612450883

Richland, L. E., Kornell, N., & Kao, S. L. (2009). Leave no test behind: Failed retrieval on a pretest can enhance future learning. *Journal of Experimental Psychology: Applied, 15,* 243–257.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology, 94,* 249–273.

Richland, L. E., & McDonough, I. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology, 35,* 28–43. http://dx.doi.org/10.1016/j.cedpsych.2009.09.001

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Mathematics: Cognitive supports for analogies in the mathematics classroom. *Science, 316,* 1128–1129. http://dx.doi.org/10.1126/science.1142103

Rittle-Johnson, B., & Schneider, M. (in press). Developing conceptual and procedural knowledge of mathematics. In R. Cohen Kadosh & A. Dowker (Eds.), *Oxford handbook of numerical cognition*. Oxford, UK: Oxford Press.

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. http://dx.doi.org/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. http://dx.doi.org/10.1037/a0014224

Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x

Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 456–468. http://dx.doi.org/10.1037/0278-7393.15.3.456

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16,* 475–522. http://dx.doi.org/10.1207/s1532690xci1604_4

Shimizu, Y. (2003). Problem solving as a vehicle for teaching mathematics: A Japanese perspective. *Teaching Mathematics Through Problem Solving: Prekindergarten-Grade, 6,* 205–214.

Smith, J. P., III, diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences, 3,* 115–163.

Star, J. R., & Rittle-Johnson, B. (2009). It pays to compare: An experimental study on computational estimation. *Journal of Experimental Child Psychology, 102,* 408–426. http://dx.doi.org/10.1016/j.jecp.2008.11.004

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning, 10,* 313–340. http://dx.doi.org/10.1080/10986060802229675

Vamvakoussi, X., & Vosniadou, S. (2012). Bridging the gap between the dense and the discrete: The number line and the “rubber line” bridging analogy. *Mathematical Thinking and Learning, 14,* 265–284. http://dx.doi.org/10.1080/10986065.2012.717378

Vosniadou, S. (2013). Conceptual change in learning and instruction: The framework theory approach. In S. Vosniadou (Ed.), *The international handbook of conceptual change* (2nd ed., pp. 11–30). New York, NY: Routledge. http://dx.doi.org/10.4324/9780203154472.ch1

Waltz, J. A., Lau, A., Grewal, S. K., Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory and Cognition, 28,* 1205–1212.

Zook, K. B. (1991). Effects of analogical processes on learning and misrepresentation. *Educational Psychology Review, 3,* 41–72. http://dx.doi.org/10.1007/BF01323662

Appendix A

Transcript Sample of the Video-Lesson

(regular type = "suggested speech", bold type = "suggested teacher actions")

Let's go back to our original problem and pull this altogether. I'd like us to think about how these strategies are different and how they are similar. This is really important part of what we are doing today.

Let's start with reviewing why Ryan's strategy is not the best way to solve this problem. Remember in Ryan's strategy, he tried subtracting "total shots tried" from "shots made" and tried to compare the missed shots **(point to the subtraction results of 8 and 9)** to figure out who was the better free-throw shooter. But we found out that this strategy does not work. Remember when Ken made 0 shots in the counterexample **(point to the counterexample),** but still missed less shots? From this example, we learned that we cannot subtract shots tried from shots made (**point to the subtraction results of 8 and 9**) and then compare the shots missed.

Subtraction is not the right way to solve this problem.

Now, how is this different from Carina's strategy?

To find out who is better she first set up Ken and Yoko's shots made and shots tried as a fraction.

Without looking at the numbers **(cover the numbers with your palm)**, how is comparing the fractions of shots made and shots tried for Ken and Yoko, in Carina's strategy (**point to the shots made and shots tried ratio)** fundamentally different from comparing shots missed only in Ryan's way **(point to the subtraction results of 8 and 9)**?

Brief Pause****

Well, Ryan only compared one unit, shots missed **(point to shots missed)**, whereas Carina compared two units **(shots made and shots tried).**

Why did she do that? Because they shot a different amount. So, if we want to know who is better at shooting free throws when they do not shoot the same number of shots, we have to compare the number of shots made and the number of shots tried.

This relationship of comparing *shots made to shots tried* is called a RATIO.

Thus,

WRITE: A relationship between two quantities is a RATIO.

So, after Carina set it up as a ratio, she made the *shots tried* of Ken and Yoko, the 20 and the 25, equal to each other. She did this by finding the LCM of, 20 and 25, which was 100 (**point to the ratio of 64/100 and 60/100)**, and then she multiplied 12 by 5 and 16 by 4 to get 60 and 64 respectively **(point to the part where Carina did the calculations on the board).** Remember, she multiplied 12 by 5 because that's the number of times she had to multiply 20 to make 100, and she multiplied 16 by 4 because that's the number of times it takes 25 to make 100. Therefore, since we found the LCM and now the *shots tried* for both Ken and Yoko are equal (***point to shots tried***), we can compare their *shots made* **(*point to shots made*).** So the point here is that we have to make the shots made equal in order to compare who is better.

This was a good strategy, but the problem with this strategy was when we tried to find the LCM for harder numbers like 19 and 25 (**point to these numbers**) we had a hard time.

We found out from Maddie's strategy that we could just divide shots made with shots tried. Let's try to figure out why Maddie's strategy works by comparing it to Carina's. This is the really important part of what we are doing today.

Something that is similar between Carina's and Maddie's, which is different from Ryan's strategy, is that they both take into account two labels: shots tried and shots made. So they use the same units to compare who is better . . .

(*Appendices continue*)

Appendix B

Problems Used in the Immediate Posttest

Construct Type	Items	Scoring and α scale
Procedural Knowledge		**Posttest α = .89** Pretest α = .86

Construct Type: Produce correct Procedures: Familiar

Problem 1.

In Cambridge, Sue and Joan played in a free-throw tournament. The results of their shooting are shown in the table below. Who is the better free throw shooter?

	Shots Made	Total Shots Tried
Sue	7	11
Joan	11	15

Please show all your work.

Who is better? _____________

Scoring:

Setup:
1 point for producing a correct solution strategy (e.g. division between 7÷11 and 11÷15)

Contender:
1 point for selecting the correct contender, if students produced a correct strategy. This ensured that students chose the correct contender by using a correct solution method, and not by chance.

Problem 2.

In Nashville, Miguel and Amos played in a free-throw tournament. The results of their shooting are shown in the table below. This time, please use TWO different ways to find who is the better free throw shooter.

	Shots Made	Total Shots Tried
Miguel	9	15
Amos	16	25

Please show all your work.

Part 1)

Way 1. Who is better ___________?

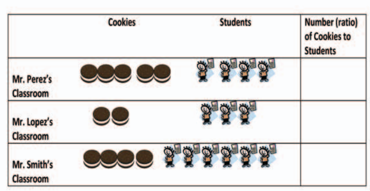
Way 2. Who is better ___________?

Scoring:

Setup:
1 point for producing at least one correct solution strategy.
OR
½ of a point when producing a correct strategy and using subtraction as an alternative strategy. This score accounted for students who believed subtraction (a common misconception) was one of the correct strategies.

Contender:
1 point for selecting at least one correct

(*Appendices continue*)

		contender and producing at least one correct strategy. OR ½ of a point when selecting a correct contender, producing a correct and 1 incorrect strategy (subtraction).
Produce correct Procedures: Transfer	***Problem 3.*** Joe and Kai played video games at boomers and then went to turn in their tickets for prizes. For every game they won, they got 1 ticket. Joe played 27 games and won 11 tickets. Kai played 11 games and won 4 tickets. Who is a better video game player? *Please show all your work.*	Setup: 1 point for producing the correct strategy. Contender: 1 point for selecting the correct contender and using a correct strategy.
Produce correct Procedures: Transfer	***Problem 4.*** Mr. Perez, Mr. Lopez, and Mr. Smith are giving out cookies to their students. The table below shows the number of cookies to students in each classroom. a) Write in the number (ratio) of cookies to students in each classroom.  b) Which two classrooms have the same amount (ratio) of cookies to students?	Omitted from analyses, because the problem was not sensitive to the intervention. Students' scores did not differ significantly between pretest and posttest (average change pre- to posttest of -5%) a) 1 point when correct for all three teachers. 0.667 of a point when correct for two teachers. 0.334 of a point when correct for one teacher.
Identify correct procedures on three sub-problems (parts), each scored on this evaluation	Is this a correct way to solve this problem? a) This is a correct way to solve this problem. b) This is a right way to solve it but the wrong answer c) No, this is NOT a correct way to solve this problem	

(*Appendices continue*)

question: d) I don't know

(Identify subtraction as an incorrect solution)

Problem 5
Part 1)

Yoko and Ken shot several free-throws in their basketball games. The result of their shooting is shown in the table.

One point for choice "c"

	Shots Made	Total Shots Tried
Ken	7	11
Yoko	3	5

Chloe solved it this way:
Ken: 11-7 = 4 missed shots
Yoko : 5 - 3 = 2 missed shots

Now that Chloe found who missed more, she compared only the shots missed, and decided that Yoko was better because she missed less shots

(Identify LCM as a correct solution strategy)

Part 2)

One point for choice "a"

	Correct Guesses	Coin Tossed
Jess	15	20
Charlie	18	25

Steven tried to solve it this way:
First he found the least common multiple for the denominators:

Jess: $\frac{\quad}{100}$ and Charlie: $\frac{\quad}{100}$

Then he found the numerators:

Jess: $\frac{7}{100}$ and Charlie: $\frac{7}{100}$

Then he compared the two fractions and decided Jess was better at guessing.

(*Appendices continue*)

(Identify division as a correct solution)

Part 3)

	Tickets Won	Games Played
Joe	11	27
Kai	4	9

Kevin tried to solve it this way:
First he divided:
Joe: 11 ÷27
Kai: 4 ÷ 9
and found that: Joe: 0 .407 and Kai: 0.44

One point for choice "a"

Procedural Flexibility Construct

Posttest α = .67
Pretest α = .57

Produce and evaluate more than one strategy for a problem

Problem 2
Please show your work.
Way 1
Way 2

Who is better __________?
Who is better__________?

Part 2)
Which way is better?
a) Way1 b) Way2

Why? a) *Less steps b) More steps but easier, c) I don't know d) It's the only way I know*

Part 1)
Setup:
1 point for producing two different correct strategies.

Contender:
1 point for selecting two correct contenders and producing two correct strategies.

Part 2)
1 point if choice "way1/way2" referred to division and the student produced at least one correct strategy, but not subtraction.

1 point if choice a) describes division as having "less steps" given the student produced two correct strategies.

(*Appendices continue*)

Identify the most efficient solution method: familiar	***Problem 6*** A weather channel in California (**TWC**) and a weather channel in New York (**KTL**) both tried to predict all the rainy days last month. The California weather channel (**TWC)** correctly predicted 5 rainy days out of 8 rainy days total. The New York weather channel (**KTL**) correctly predicted 14 rainy days out of 21 total rainy days. Which strategy will tell us which channel was more **accurate**, in the **least** number of steps: ***Circle your answer:*** *a) Divide 5÷8 and 14÷21* *b) Multiply 5*8 and 14*21* *c) Find the least common multiple for 21 and 8* *d) Subtract 8- 5 and 21-14*	1 point for choice a)
Identify the most efficient solution method: Transfer	***Problem 5 Part 3)*** was used with the following question: After thinking about it, Steven realized that he could also find out who is better by finding the least common multiple for the numerators at the start of the problem: Charlie: $\frac{90}{}$ Jess: $\frac{90}{}$ Is this a correct way to solve this problem? a) This is a correct way to solve this problem. b) This is a right way to solve it but the wrong answer c) No, this is NOT a correct way to solve this problem d) I don't know	1 point for choice a)

(*Appendices continue*)

Conceptual Understanding Construct		**Posttest α=.66** Pretest α = .42
Adapting procedure to a novel problem type and context	***Problem 7***	1 point for producing a correct solution strategy 1 point for selecting the correct contender, but only if they produced a correct strategy

Adelina's Lemonade	**Marcos' Lemonade**
	

Whose lemonade tastes more "lemony?"

____

Show all your work.

Recognizing the correct solution & setup on novel problem type and context:	***Problem 8.*** Yoko decided to divide her cookie jar amongst her friends.	Problem omitted due to ceiling effects (averages ranged between 81%-89%) and they were not sensitive to the intervention (average change pre- to posttest -3%) Part 1) 1 point for choice a) Part 2)

	Cookies	Friends
Yoko		

Part 1) To figure out how many cookies each friend gets, how should you set up the problem?

Circle your answer

a. 

b. 

c. 

d. 

(Appendices continue)

	Part 2) What units go with your answer in Part 1? a) cookies per friend b) friends times cookies c) cookies d) friends	1 point for choice a)
Evaluating explanation for solution strategy	***Problem 1 (see above)*** was followed by the following question: How do you know? *a) I compared the number of shots made* *b) I compared the number of shots tried* *c) I compared the shots made to shots tried* *d) I compared the number of shots missed*	1 point for choice c) only if student used a correct strategy and selected the correct contender for problem 1.
Knowledge of units:	***Problem 5 Part 2) and Part 3)*** were used with the following questions: Write the labels that go with these numbers:	
	Part 2) Jess: $\frac{75}{100}$----------- Charlie: $\frac{72}{100}$ -------------	Part 2) 1 point for writing "$\frac{\text{coins tossed}}{\text{correct guesses}}$"
	Part 3) Joe: 0 .407 ---------------------- Kai: 0.44 ---------------------	Part 3) Omitted from analysis; Only 6 out of 97 students responded correctly. 1 point for writing "$\frac{\text{games played}}{\text{tickets won}}$"
	Problem 5 Part 3) was used with the following question: What do the numbers .407 and .44 represent? (*Circle your answer*) a) *The number of games played for each ticket won* b) *The number of tickets won* c) *The number of games played* d) *The number of tickets won for each game played*	1 point for choice a)

(*Appendices continue*)

	Problem 4 part b) was used. a) Which two classrooms have the same amount (ratio) of cookies to students?	Omitted from analyses (see Problem 4) 1 point for correct answer.
Produced Misconception *Used strategy shown to be invalid during instruction*		**Posttest α = .66** Pretest α = .46
	Problems 1, 2, 3, and 7	1 point if students used subtraction as a solution strategy.

See the online article for the color version of this appendix.

Received October 28, 2014
Revision received April 29, 2015
Accepted May 1, 2015 ■