ARTICLE IN PRESS

Contemporary Educational Psychology xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Learning by analogy: Discriminating between potential analogs

Lindsey E. Richland [a,*], Ian M. McDonough [b]

[a] University of California, Irvine, Department of Education, United States
[b] University of Chicago, Department of Psychology, United States

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Cognition
Analogy
Comparison
Problem solving

ABSTRACT

The ability to successfully discriminate between multiple potentially relevant source analogs when solving new problems is crucial to proficiency in a mathematics domain. Experimental findings in two different mathematical contexts demonstrate that providing cues to support comparative reasoning during an initial instructional analogy, relative to teaching the same analogs and solution strategies without such cues, led to increased ability to discriminate between relevant analogs at a later test. Specifically, providing comparative gestures and visibly aligned source and target problems during initial learning led to higher rates of positive extension of learning to new contexts, and lower rates of susceptibility to misleading contextual features, both immediately and after a week delay.

1. Introduction

Instructional analogies provide opportunities for teachers to clarify similarities and differences among problems, concepts, and procedures; or among misconceptions and correct strategies (see Brown & Clement, 1989; Chen & Klahr, 1999; Clement, 1993; Duit, 1991; English, 2004; Gentner, Loewenstein, & Thompson, 2003; Kolodner, 1997). Such analogies may compare a mathematical representation to a non-math entity (e.g., balancing equations is like balancing a scale), but they may also compare two math concepts or methods. For example, a teacher might compare percent and ratio problems, demonstrating their common structure. While experts easily identify such concepts as similar, student learners may miss such connections unless they are very explicitly taught. Such comparisons form a backbone of mathematical thinking and learning, enabling students to build understanding of new topics based on their prior knowledge.

Experimental studies have demonstrated that encouraging students to draw such connections facilitates learners' ability to use the concepts or methods in a future context (e.g., Rittle-Johnson & Star, 2007). However, students must also be able to regulate and resist over-extensions of potential analogies. For instance in the above example, one cannot extrapolate from the ability to add percentages ($10\% + 5\% = 15\%$) to adding ratios ($3{:}5 + 3{:}5 = 6{:}10$). Accordingly, students must be able to identify and differentiate the conceptual structure of mathematical representations in order to identify relationships with previously learned principles from memory (Bransford, Brown, & Cocking, 2000; National Mathematics Advisory Panel, 2008a, 2008b).

While essential, learners often show difficulties in noticing commonalities between previously learned and new mathematics across contexts (see Bassok & Holyoak, 1993; National Research Council, 2001; Novick & Holyoak, 1991). In the classroom, learners' failure to recognize word problems as comprising known concepts was recently cited by teachers as one of the most pressing problems in teaching algebra (National Mathematics Advisory Panel, 2008a, 2008b). Word problems and equations may appear different at a surface level, even though the underlying mathematical structure is common.

The literature in analogical reasoning makes a distinction between correspondences based upon surface features, or appearance (e.g., a word problem about trains) and those based on deep structure (e.g., mathematical structure) (Gentner, 1983). In a mathematical instructional context, therefore, there are at least two levels at which mathematics problems can share correspondences with previously instructed problems. New problems may or may not share mathematical structure such that they can be solved in similar ways. In addition, they may or may not appear similar at a surface level independent of mathematical structure. Thus, problems that are mathematically similar may *appear* different (e.g., two word problems appear different if one is about pizza and the other about dividing work hours, even if they are mathematically equivalent), and problems that are mathematically different may *appear* similar (e.g., two word problems about trains that are mathematically dissimilar).

This distinction has proven useful in understanding the challenges for mathematics students. When contexts are less well understood, novices tend to map correspondences based on surface features (e.g., Reed, 1987; Ross & Kennedy, 1990; Schoenfeld

* Corresponding author. Address: University of California, Irvine, 2001 Berkeley Place, Irvine, CA 92697, United States.
E-mail address: lerich@uci.edu (L.E. Richland).

0361-476X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.cedpsych.2009.09.001

Please cite this article in press as: Richland, L. E., & McDonough, I. M. Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology* (2009), doi:10.1016/j.cedpsych.2009.09.001

& Herrmann, 1982), while they tend to map better-understood contexts based on deep structure (e.g., Chi, Feltovich, & Glaser, 1981; Chi & Ohlsson, 2005, chap. 14; Novick, 1988; Schoenfeld & Herrmann, 1982). Mapping correspondences based on mathematical structure leads to more flexible, expert-like mathematical proficiency. Thus, seeking instructional strategies for improving learners' ability to spontaneously notice and attend to structural correspondences is a crucial research goal.

2. Cues to draw attention to relational similarity

The current study experimentally tests an instructional strategy for improving learners' likelihood of noticing and learning from instructional analogies. Specifically, the study uses videotaped instruction in a laboratory setting to determine whether cues to support learning from a presented comparison would impact students' later rates of positive and negative extension on a posttest.

The tested instructional strategy derived from two sources. First, the basic literature on analogical reasoning suggests that explicit cuing and reducing processing demands should improve learning. Second, a video analysis of typical 8th grade mathematics instruction in the US and two higher achieving countries, China (Hong Kong) and Japan, identified classroom-relevant strategies for cuing and reducing processing demands, and revealed that teachers in both the higher achieving countries provided more support for learning from analogies than did the US teachers. This background is briefly reviewed below, and then the intervention is described more specifically.

2.1. Basic studies of analogy

One line of well-replicated studies has shown that providing cues at the time of final test improves the likelihood that learners notice the relations to prior instructed problems (Gick & Holyoak, 1980). These cues may take the form of general reminding statements that encourage participants to consider prior instruction, or of more specific information about which elements might be mapped successfully (e.g., Blessing & Ross, 1996; Novick, 1988; Novick & Holyoak, 1991).

A second strategy has involved providing multiple analogs during instruction and encouraging learners to map the structural correspondences as a learning tool, rather than simply as a final test. The act of mapping relational correspondences between problem analogs, based on analogical reasoning, seems to operate as a strong learning opportunity, leading to increased rates of spontaneous transfer to new problems at the time of a final test (Catrambone & Holyoak, 1989; Gick & Holyoak, 1983; Novick & Holyoak, 1991; Rittle-Johnson & Star, 2007; Ross & Kennedy, 1990).

In the classic demonstration of this effect with general insight problems, Gick and Holyoak (1983) showed learning gains for providing two analogous problems (one about a fireman, one about a General attacking a castle) during a learning phase and giving participants questions that asked them to map correspondences between elements in the two problems (e.g., "what is like the General in the fireman story?"). In the mathematics domain, studies by Novick and Holyoak (1991), Ross and Kennedy (1990) revealed that giving reminder hints during an initial problem solving opportunity led to higher performance on a final test for those who successfully performed the analogical mapping in the first instance.

Rittle-Johnson and Star (2007) recently showed benefits of having pairs of students compare two solution strategies (one standard, one non-standard) to distributive property problems during learning. At a final test, those participants significantly outperformed other student pairs who saw the same problems and solutions but on different pages. This result was obtained on final tests of retention as well as using measures of schematic, conceptual understanding.

Despite these many studies finding benefits of analogical instruction, providing an instructional analogy far from guarantees spontaneous use of the analogs on a further test. Several studies have shown little benefit for various strategies designed to encourage learners to map correspondences during instruction. Reed (1989) found no benefits for providing two algebra analogs during instruction and prompting comparison using a paper–pencil worksheet packet. He argued that this failure may have been due to the high domain knowledge required in algebra learning. Similarly Gerjets, Scheiter, and Catrambone (2006) found no benefits of having learners generate self-explanations during analogical learning for molar and modular worked examples (problems taught with focus on the overall solution strategy or broken into smaller, more easily processed meaningful components, respectively). In related work, Scheiter, Gerjets, and Catrambone (2006) did find that static pictures and cues requesting learners to visually imagine the steps of worked examples facilitated learning, but this was not the case when dynamic animations were provided. The authors suggested that higher processing load was necessary to interpret and remember what had been shown in the animation since it was no longer available (whereas the crucial steps were available when presented in static photographs). By contrast, reducing cognitive load by breaking problem analogs into subgoals and chunking steps into labeled groupings facilitated learning and transfer (Catrambone, 1998).

Many of the existing studies seem to coalesce on this point: higher cognitive load reduces learners' ability to focus on structural commonalities during instruction, which makes them less likely to benefit from the instructional analogy. Instead, learners tend to focus on surface similarities when under load (e.g., "Ah, this problem is about pizzas. I just have to remember how we solved that last pizza problem"). Basic studies of analogical reasoning support this view. Adding a working memory load when undergraduates solved picture analogy problems led to greater rates of mapping correspondences based on surface features rather than on structural correspondences (Tohill & Holyoak, 2000; Waltz, Lau, Grewal, & Holyoak, 2000). Increasing working memory load by making source and target analogs not visible together, as in the study of Rittle-Johnson and Star (2007), also made participants more susceptible to distraction from surface features and to the complexity demands of the analogs themselves (Cho, Holyoak, & Cannon, 2007). The cognitive demands of analog complexity and distraction from irrelevant surface features are particularly striking for those with already somewhat limited cognitive resources, such as young children (Richland, Morrison, & Holyoak, 2006) and aging adults (Viskontas, Morrison, Holyoak, Hummel, & Knowlton, 2004).

Thus instructional analogies may be most effective when taught in a way that reduces processing load for learners as much as possible. But a question remains. What exactly does this mean in the context of realistic, everyday mathematics classroom instruction?

Recent practice recommendations (Pashler et al., 2007) have derived from a body of work informed by Cognitive Load Theory (Sweller & Cooper, 1985). Cognitive Load Theory builds on the information processing model of memory and in particular, limits in working memory capacity, to argue that instructional designers should consider the inherent demands of instructional materials and tasks so as not to overwhelm learners. Importantly, these constraints are particularly important for novices in a domain (e.g., Catrambone, 1998; Paas & Vanmerriënboer, 1994; Renkl & Atkinson, 2003; Sweller, Chandler, Tierney, & Cooper, 1990; Zhu & Simon, 1987). Optimal learning environments may be quite different for experts versus novices, based on experts' ability to schematize, or group, information while novices must instead interact with all relevant problem solving components separately (Kalyuga, Ayres, Chandler, & Sweller, 2003; Mayer, 2001).

Please cite this article in press as: Richland, L. E., & McDonough, I. M. Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology* (2009), doi:10.1016/j.cedpsych.2009.09.001

Together, these laboratory-based research programs provide a theoretical explanation for why learners require additional support for learning from comparisons between analogs. Laboratory studies provide some insight into those supports – opportunities to study multiple analogs simultaneously with explicit visual or verbal cues to reduce the cognitive processing load for novices. Next we discuss a classroom study that used these principles to identify everyday classroom practices that seemed likely to support students' analogical thinking.

2.2. Classroom-based strategies for supporting comparisons

Richland, Zur, and Holyoak (2007) examined typical classroom analogy practices in 8th grade mathematics classrooms videotaped as part of the Trends in International Mathematics and Science Study (Hiebert et al., 2003). By comparing US teachers' practices with those of teachers in higher achieving countries, the authors sought to gain insight into classroom feasible strategies for supporting high quality learning opportunities.

Richland et al. (2007) studied US, Hong Kong, and Japanese lessons, sampled from the larger TIMSS-R dataset which was a randomized probability sample of all lessons taught in each region over the course of an 8th grade academic year. Teachers' uses of instructional analogies were analyzed within ten randomly sampled lessons from each country, each lesson taught by a different teacher. Each lesson was analyzed first for presence of instructional comparisons between the mathematical structure of a problem or concept and another problem, concept, or non-math example. Next, each such comparison was analyzed according to a set of six quantitative codes that assessed adherence to principles likely to reduce learners' processing load and enhance relational learning. These codes derived from an integration of the literature and observations of the videos.

Analyses revealed that teachers across countries used similar total numbers of comparisons between two or more analogs (between 7 and 20 per lesson). However, these patterns varied dramatically when comparisons were assessed for adherence to principles likely to reduce cognitive load. Codes measured: (1) learners' familiarity with the source, (2) whether source analogs were presented visually versus only orally, (3) the source's visibility during instruction of the target, (4) presence or absence of visuo-spatial cues to the correspondences between items being compared, (5) presence or absence of gestures that move back and forth between the items being compared, and (6) the use of mental imagery or visualizations. These strategies presumably supported learner's processing by increasing their relative domain expertise (using familiar sources and mental imagery), reducing demands on working memory (source presented visually, left visible during target instruction), and reducing demands on attention by drawing learners' eyes to the comparison itself (comparative gesture and visuo-spatial cues within visual representations).

The findings were quite clear: for all codes, the US teachers were least likely to use these cuing support strategies. Teachers in Hong Kong and Japan were significantly more likely to use all of these strategies, sometimes using them in double or triple the percent of instructional analogies.

Richland et al.'s (2007) data indicate that these cuing support strategies are not yet being used regularly by US teachers, and could provide new ways to optimize US teachers' current uses of instructional analogies. A limit to this study, however, was that no student outcome data were available from the TIMSS studies to directly tie instructional practices to student learning (Hiebert et al., 2003; Perry, Vanderstoep, & Yu, 1993). Thus while there is theoretical support for these strategies as methods for increasing students' ability to learn from relational correspondences, a more direct test is necessary.

3. Experimental test of cuing during instructional analogy

The current study uses a controlled, laboratory format to provide a direct test of the utility of these strategies for improving students' learning from a mathematics instructional analogy. Two experiments were conducted comparing undergraduates' learning from analogies in which the most common of these strategies were used, versus learning from the same content but in which these supports were not used. The instruction in each case used a videotaped lesson to simulate whole-class instruction and to maximize comparability between conditions. Highly supported analogies involved a structural comparison in which the teacher used a combination of the four most common strategies identified in the coding study (source presented visually, source remains available during target analog instruction, visual alignment between the two analogs, and gestures between the two analogs). Minimally supported analogies involved instruction using the same two analogs but without any of the above support cues.

The two experiments were conducted in different mathematical contexts to maximize the generalizability of the results. Experiment 1 provided instruction in permutation and combination problems, and Experiment 2 involved teaching a limit to the linearity assumption of proportional reasoning. These materials were selected for several reasons. Both are mathematical content areas in which undergraduates are well known to demonstrate a lack of success, but both are included on the GRE and so are at a suitable level for undergraduate students. Also several analogical studies have examined the permutation, combination context (e.g., Ross, 1989; Ross & Kennedy, 1990; Vanderstoep & Seifert, 1993) so we can build on prior work in Experiment 1, and expand to a new context in Experiment 2. Finally, we sought to generalize across the two most common types of mathematical analogies in US instruction (Richland, Holyoak, & Stigler, 2004), between two problem types (Experiment 1) and between two solution strategies (Experiment 2).

Posttests were administered immediately (Experiments 1 and 2) and after a week delay (Experiment 2). Posttests measured learners' ability to solve test problems that: (a) appeared similar to instructed problems, and (b) in which the appearance and mathematical correspondences were cross-mapped. Cross-mapping meant here that the story context for one instructed problem was the same as the story context for a posttest problem with dissimilar mathematical structure. The posttest measures derived from the complexities of adequately measuring analogical learning. The literature on such measurements is briefly reviewed. Following this review, the experiments are reported.

4. Measuring analogical learning

Several strategies have been used in the literature to measure analogical learning. Novick and Holyoak (1991) use the term "analogical transfer" to describe the retrieval of a previously taught problem to help generate potential solutions to a test problem. Cross-mapping surface and structural similarities on test problems has also been used to assess the depth of learners' understanding and their focus on structural relations, as well as their resistance to irrelevant surface cues (e.g., Gentner & Toupin, 1986; Goldstone & Medin, 1994; Novick, 1988; Ross, 1987, 1989; Ross & Kilbane, 1997).

In a paradigm closely related to the current Experiment 1, Vanderstoep and Seifert (1993) taught participants how to solve two mathematically similar (permutation and combination) problems, and then tested participants' ability to determine which of the instructed procedures would be most appropriately applied to test problems. The authors drew on Schwartz and Bransford's (1998) notion of contrasting cases, comparing analogs that shared high structural similarity aside from a key variation. To measure learning, the authors distinguished between learning *how* and learning *when* to use solution strategies, or in Ross's (1989) terms, retrieval access and use.

Vanderstoep and Seifert's (1993) experiments support the conclusion that considerable instructional support is necessary to allow learners to benefit from a comparison between similar or contrasting analogs. Written text-based instructional manipulations tested the benefits of an explicit rule for when to use each formula versus providing worked solutions to the two problems separately or together. The authors measured two indicators of learning: (1) accurate problem solving and (2) ability to determine which instructed problem solution matched each test problem. Findings showed no differences between conditions in solution accuracy, but providing the explicit rule for when to use each formula led to improvements in appropriate formula use. Allowing participants to induce the rule by providing the two worked problems together did not improve solution selection.

These experiments point out the importance of experimental design and posttest measures that allow for disentangling the effects of teaching on (1) memory for instructed problems and solutions and (2) ability to know when to use these solutions. These types of knowledge seem to be distinct, such that one might recall a solution strategy but not spontaneously notice the utility of that strategy at the time of test.

Posttest measures in the current experiments therefore were designed to provide data on both points through a manipulation of surface and structural similarity. A cross-mapping procedure was used to create problems in which the setting of a problem taught during instruction either mapped directly onto the setting of a problem solved the same way on a posttest problem (facilitory similarity), or mapped onto the setting of a problem solved a different way (misleading similarity). This measurement decision allowed us to begin to differentiate between alternative characterizations of the impact of cuing on learners' later problem solving. Four alternative explanations were possible for how cuing might aid learning. Increased cues during analogical instruction might lead to: (1) abstracted schematic representations of the two analogs, (2) production of a decision rule for when to apply each taught solution strategy, (3) improved retention of the individual problems, or (4) more expert-like processing of problems (attention to structural versus surface features). These possibilities were explored in the data gathered in the following two experiments. Results from the two mathematical contexts are discussed separately, and then the strength of generalizability across the two contexts is used in a final discussion to try and determine (a) whether providing support in the form of analogy cues improves learning, and if so, (b) disentangle these four possibilities for how the learning is improved.

5. Experiment 1

Experiment 1 compared undergraduates' learning from a highly versus minimally cued instructional analogy between a permutation problem and a combination problem. We hypothesized that the highly cued instructional analogy would lead to greater ability to map instructed problems to test problems on the basis of structure.

5.1. Method

5.1.1. Participants

Seventy undergraduates participated for partial course credit (56 females, $M = 21$ years). Data from six of these participants were excluded for failure to complete the study. Participants were recruited from the university subject pool; generally these were students taking introductory psychology and education courses. For a baseline measurement, 28 additional undergraduates (23 females, $M = 21$ years) were given the posttest only.

5.2. Materials

5.2.1. Instructional videos

Videotaped instruction taught students about two math concepts tested on the GRE: permutations and combinations. Permutation problems involve determining the maximum possible arrangements of a set of items (e.g., the number of alternative orders that four runners can be awarded a gold and silver medal). Combinations are permutation problems with the additional constraint that certain alternative orders of the same objects are considered the same (e.g., the number of different ways that four students can be awarded two entrance tickets to a show, where receiving ticket number one and ticket number two is functionally the same). The problem solution therefore must exclude equivalent orders.

Videotaped instruction was developed for two conditions: (a) high cuing for comparison and (b) minimal cuing for comparison. Separate videos were constructed for the two conditions, allowing for instruction that simulated classroom teaching but was easily controlled between conditions. Both videos taught the same two problems and solution strategies, and were approximately the same length (11.5 and 13 min, respectively). In both videos, the learner first received instruction about solving permutation problems embedded within an example of runners in a race. Next, they received instruction about solving combination problems embedded within an example of students vying for tickets to a lecture. The key instructional elements of the two videos are described and aligned in Appendix A.

The instructional manipulation involved the manner in which support was provided for the learners to notice and benefit from a comparison between the source and target analogs. The high cuing video included explicit cues to the alignment and correspondences between the two analogs. Instruction in this video included the four most common cuing and processing support strategies identified within the cross-cultural video analysis: comparative gesture, source visually supported, source visible during instruction of target, and visual alignment. The teacher gave the same instruction about permutations provided in the minimal cuing condition, but in the high cuing video, the teacher left the problem on the board while he described how to solve the combination problem. Learners could still visually reference the permutation problem while they were learning about the combination problem to check commonalities or differences. The teacher in the high cuing video also used visual supports to promote the students' comparisons between these two problems. He gestured back and forth between the two problems (comparative gesture) and wrote the problems on the board in a linear, sequential style that highlighted the alignment between the two. For example, each line of the problem solution contained an underscore for numbers not yet filled in. This procedure ensured that the structure was preserved for every line of the problem (see Fig. 1), and there was a record of the sequence of the steps visible on the board.

By contrast, in the minimal cuing video, the commonalities between the two problems were not immediately apparent by looking at the spatial information on the board (see Fig. 2). There were two main differences. First, the text of the two problems were each written out more completely than in the high cuing condition, which did not as explicitly highlight the key problem elements. The solution methods were also written differently, such that there was not a new sequence of underscore lines written for each step of the

Please cite this article in press as: Richland, L. E., & McDonough, I. M. Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology* (2009), doi:10.1016/j.cedpsych.2009.09.001

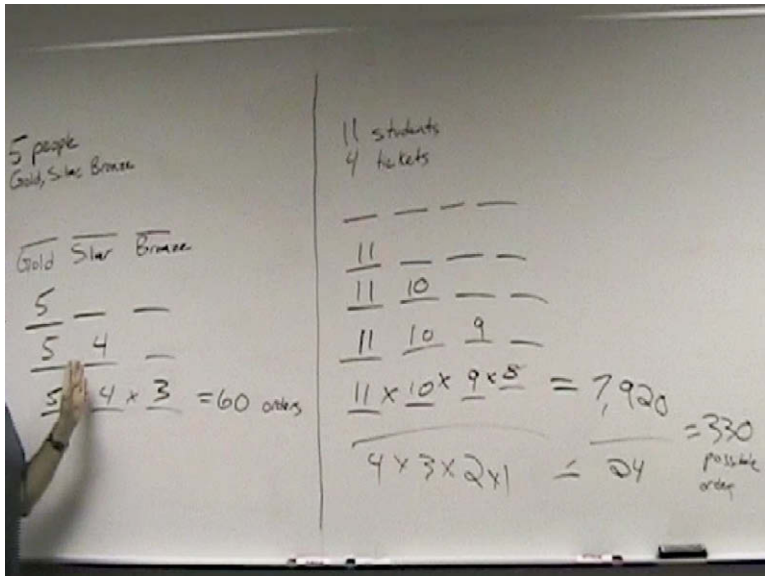


Fig. 1. Final screen from Experiment 1 instructional video: high cuing comparison.

problem. Rather, the underscore lines were only written once, and the steps were all completed using that same row of lines. This was more efficient than in the high-cue condition, and is the way one would typically solve the problem independently. However, it did not provide the same visual record and required that the viewer recall in working memory the sequence of the steps.

5.2.2. Posttest

Two kinds of key test problems were developed for the posttest, resulting in four total problems used in analyses. Two *Facilitory Similarity* problems were designed to align with both the mathematical and the surface story context features of the instructed problems. One problem was a permutation problem about a race, and the second was a combination problem about tickets to a lecture.

The second key test problems were *Misleading Similarity* problems, in which the word problem context was misaligned, or cross-mapped, with the mathematical structure. The misleading similarity permutation problem was set in the context of procuring tickets to a lecture, and the misleading similarity combination problem was set in the context of a race. These problems were presented in randomized order.

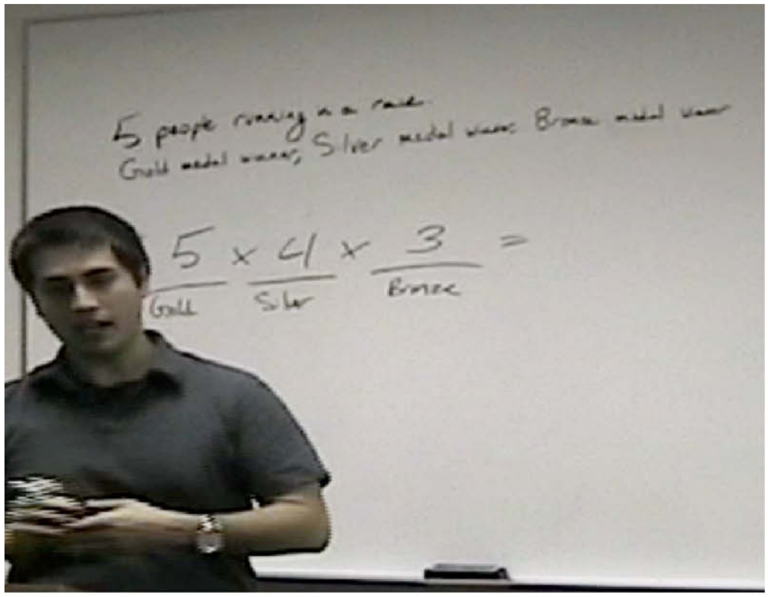
Each posttest contained a permutation problem and a combination problem with facilitory similarity and one of each with misleading similarity, yielding a total of four key test problems. Each participant also solved five additional word problems of similar length. These served as distractor problems and tests of learners' over-extensions of the solution procedures. The five problems were selected in ten balanced configurations from the following set: three factorial problems (one about races, one about tickets to lectures, and one set in an irrelevant context); and four more advanced combinatoric problems (not directly solvable with the instructed solution), either set in the context of a race or a lecture. These were interspersed in randomized order between the key, analyzed problems, in order to simulate the necessity in classroom settings for learners to identify relevant analogs amid other distracting problems.

Baseline performance on posttest problems were assessed on a separate group of students sampled from the same population as in the experimental conditions, but who did not watch any of the instructional videos. These data were collected to ensure that students would not begin the instruction at ceiling levels of performance. An independent group of participants was tested, rather than using a pretest format with the same participants who completed our study, because of the sensitivity of the instructional manipulation. A growing body of literature suggests that testing itself can provide a learning benefit for test-takers (see Roediger & Karpicke, 2006), even when there is no feedback provided and when all answers to pretest questions are incorrect (Richland, Kornell, & Kao, 2009). We were particularly concerned about this in the current experimental case, since we didn't want our participants to spontaneously compare and contrast the permutation and combination problems on the pretest. That initial experience might have directly impacted their learning from the two conditions, perhaps influencing each condition differently.

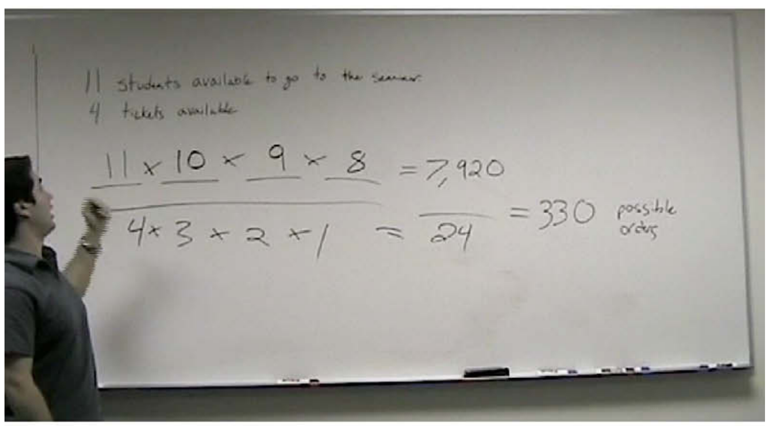
Participant responses on all test questions were scored both for their accuracy and for the type of error made. Accuracy was scored as percent of problems set up correctly. Correct set up included both use of the appropriate formula and adaptation to the correct values within the target problem context. Calculation errors were not taken into account in accuracy scores, since our goal was to measure appropriate decisions about how to classify problems and map them onto source analogs. Partial credit for retrieving the source analog without ability to draw inferences about the target problem (adaptation) was not assigned, since the design of the posttest made retrieval without understanding uninteresting. Correct set up was scored as 1; all other answers were scored as 0.

Second, instances of cross-mapping were assessed, since these were diagnostic of failures to appropriately discriminate between analogs on the basis of structure. Errors were coded as cross-mapping errors when participants used a combination strategy on a permutation problem and vice versa. Other types of solution errors were not coded as cross-mapping attempts. All participants who attempted to use the strategy that would be correct for the alternative type of problem (i.e., using a permutation strategy to solve a combination problem or vice versa) were given a 1, while those who scored correctly or made a different type of error were scored 0.

The distractor problems were coded by examining whether the participants made expected errors, which meant using correctly either the permutation or combination solution to that test problem, though neither was actually a valid way to answer the test question. This was the same as the way the cross-mapping scores were calculated, but we scored both whether participants made an attempt to use a combination or a permutation solution. As before, calculation errors were ignored in this scoring.



(a) Source Analog



(b) Target Analog

Fig. 2. Experiment 1 instructional video: low cuing comparison.

5.2.3. Procedure

In a between-subjects design, participants were randomly assigned to one of the two instruction conditions (minimal cuing condition, $N = 30$; high cuing condition, $N = 34$). Participants were tested individually with a trained experimenter in a laboratory setting. They entered and were seated in front of a Macintosh computer. They were told they would watch a short mathematics lesson, and were shown either the high or low cuing video. They were then given a 5-min timed word-finding distractor task. Finally, participants were given a paper packet that included all of the posttest problems and distractor problems, with two problems per page. This final testing was not timed. Participants were not given any explicit cues to use the videotaped instruction when solving the posttest problems.

5.3. Results

Analyses were conducted on the two types of dependent variables: accuracy and presence of a cross-mapping error. Accuracy data are first described below, and then data from cross-mapping errors are presented.

5.3.1. Baseline accuracy

Participants who did not watch either of the instructional condition videos provided evidence that undergraduates at this university were not overall mathematically competent on the test problems prior to instruction. The four problems of interest were analyzed separately. Participants were approximately equally accurate on all four problems. For the problems that served as facilitory similarity problems in the experimental conditions, accuracy was 10% correct for the permutation problem ($SD = 30$), and 3% correct for the combination problem ($SD = 19$). For the problems that served as misleading similarity problems in the experimental conditions, accuracy was 10% correct for the permutation problem ($SD = 30$), and 10% correct for the combination problem ($SD = 30$). Thus these data show that for the studied population of undergraduates, participants were unlikely to solve any of the posttest problems correctly prior to instruction, with accuracy rates ranging from 3% to 10%.

5.3.2. Posttest accuracy

Analyses next examined the impact of the instructional manipulation on posttest performance. Two types of dependent mea-

Please cite this article in press as: Richland, L. E., & McDonough, I. M. Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology* (2009), doi:10.1016/j.cedpsych.2009.09.001

sures were included in a repeated-measures model. Performance on facilitory similarity problems was viewed as an index of participants' retention and ability to use the instructed problem solutions when minimal demands were placed on noticing the relevance of the solution strategy. Performance on the misleading similarity problems was viewed as an index of participants' ability to ignore irrelevant surface cues, and instead notice and extend the instructed solution strategies to a context that appeared misleading and dissimilar.

A repeated-measures ANOVA used accuracy as a dependent variable with two within-subjects levels of surface similarity (facilitory versus misleading similarity) to examine the effects of the between-subjects independent variable of instructional condition (high versus low cuing). As can be seen in Fig. 3, there was a main effect of surface similarity $F(1, 62) = 29.9$, $p < .001$, $\eta_p^2 = .33$, such that a higher proportion of facilitory similarity problems (problems that shared surface and mathematical similarity with the instructed problems) were set up correctly ($M = .80$, $SD = .32$) than were problems with misleading surface similarity ($M = .51$, $SD = .38$). Thus, as expected, the manipulation of surface similarity was effective and accuracy was higher when surface similarity was correlated with structural similarity than when the surface similarity was misleading.

Surface similarity also significantly interacted with instructional cuing condition, $F(1, 62) = 4.05$, $p < .05$, $\eta_p^2 = .06$. Importantly, this interaction revealed that rates for the proportion of problems set up correctly were high for both conditions on the facilitory similarity problems ($M = .80$, $SD = .31$, and $M = .81$, $SD = .32$). However, there was a difference in accuracy by condition for the misleading similarity problems. A separate univariate analysis of the misleading similarity problems revealed that the average proportion correct for the high cuing condition ($M = .62$, $SD = .36$) was significantly higher than for the low cuing condition ($M = .41$, $SD = .40$), $F(1, 62) = 4.57$, $p < .05$, $d = .55$. This finding suggests that those participants in the former condition were more attentive to structure, as opposed to surface features, and were better able to discriminate between problems on that basis. It appears that high cuing led to greater schematization of the source analogs, and thus lowered participants' susceptibility to misleading surface similarity.

Even so, these participants were evidently not reasoning *entirely* on the basis of fully schematized, or rule-based, representations. The main effect of surface similarity suggested that even the high-cue condition did not produce fully schematized knowledge. This interpretation was supported by an analysis comparing only the high-cue condition scores on the facilitory and misleading similarity problems. A paired-samples *t*-test revealed the difference remained significant, $t(29) = .26$, $p = .01$, $d = 2.69$, indicating that even following highly cued relational instruction, source analogs were retrieved as whole problems rather than simply as solution schemas or decision rules. The difference between conditions appears to reflect which elements of a target problem were used as retrieval cues for selecting between the two potential source analogs.

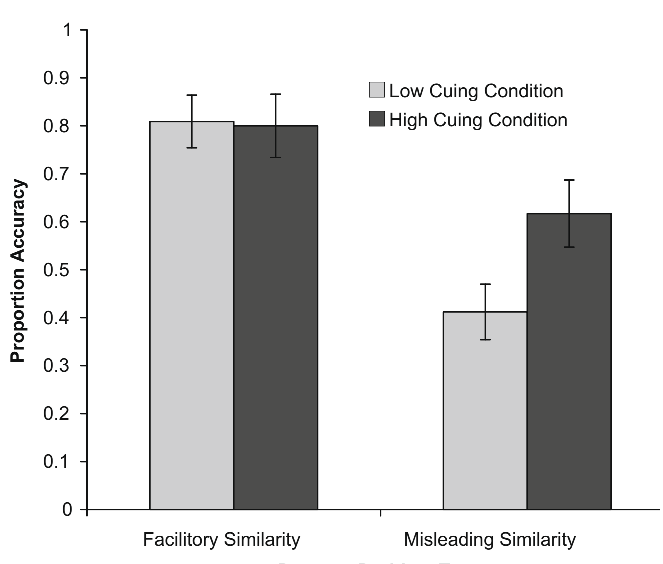


Fig. 3. Impact of support for analogical instruction on posttest accuracy (Experiment 1).

5.3.3. Cross-mapping errors

The cross-mapping error data further support the interpretation that high cuing during instruction impacted participants' later spontaneous attention to relational structure as opposed to surface features. The same analyses were repeated using cross-mapping errors as a dependent variable. These data again showed a main effect of surface similarity, $F(1, 62) = 36.38$, $p < .001$, $\eta_p^2 = .37$, supporting the validity of the surface similarity manipulation. This analysis also revealed an interaction between instructional cuing and surface similarity that approached significance, $F(1, 62) = 3.14$, $p = .08$, $\eta_p^2 = .05$. As shown in Table 1, the pattern of errors supports the interpretation drawn above for the accuracy data. Subjects in both instructional conditions made relatively few of these diagnostic errors on the facilitory similarity problems ($t(62) = 1.86$, $p = .85$, $d = .05$); but on misleading similarity problems, the low cuing condition resulted in marginally higher error rates than did the high cuing condition ($t(62) = 1.9$, $p = .06$, $d = .47$). This pattern of errors suggests that participants in the low cuing condition were more likely to retrieve a source analog on the basis of surface similarity. In contrast, those in the higher cuing condition were presumably more likely to retrieve a source analog on the basis of relational structure, and thus were less impacted by the cross-mapping procedure.

5.3.4. Over-extension to distractor problems

Distractor problems were analyzed for errors in which participants attempted to map taught solution strategies to the mathematically dissimilar problems that shared context features. There were no differences across conditions: 30% of participants in both conditions sought to map a solution strategy based on the surface context, even when the problems were not mathematically comparable. This finding suggests that when they lacked the content knowledge to interpret test problems mathematically, all participants moved to map correspondences on the basis of surface features in some instances. This result also supports the interpretation that even participants in the highly-cued comparison condition were not reasoning on the basis of fully schematized knowledge, but rather maintained contextual information in their knowledge representations.

5.4. Discussion

Experiment 1 was conducted with two main aims: to determine whether cues to support analogical thinking would increase stu-

Table 1
Frequency of cross-mapped strategy use across problem type in Experiment 1.

Instructional condition	Facilitory similarity problems	Misleading similarity problems
High cuing	7 (16)[a]	28 (31)
Low cuing	6 (17)	46 (41)

[a] Standard deviations listed in parentheses.

Please cite this article in press as: Richland, L. E., & McDonough, I. M. Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology* (2009), doi:10.1016/j.cedpsych.2009.09.001

dents' flexible learning for instructed concepts, and if so, to better understand how the cues would change learners' representations of the instruction. In answer to the first, the data support the interpretation that cuing does provide a benefit for encouraging flexible learning. Providing more analogical cues did not affect learners' *retention* of instructed solution strategies, participants were able to solve facilitory similarity problems after both videotapes, but it did impact their ability to recognize and extend the solution strategies to contexts with misleading surface appearance. Participants with higher cuing were better able to recognize problem structure and use the appropriate solution strategy on the misleading similarity problems. Participants who received minimal cuing were most likely to make cross-mapping errors on misleading similarity problems. This suggests that they were more reliant on surface similarity in determining when and where to use a learned solution strategy.

These data reiterate the importance of posttest measures that allow for disentangling the effects of teaching on memory for instructed solutions and ability to know when to use these solutions. Further, these data allow for beginning to understand why the cues were successful. Four alternative explanations were originally posited for how cuing might aid learning. Increased cues during analogical instruction might lead to: (1) abstracted schematic representations of the two analogs, (2) production of a decision rule for when to apply each taught solution strategy, (3) improved retention of the individual problems, or (4) more expert-like processing of problems (attention to structural versus surface features).

Based on Experiment 1 data, there seems to be little support for the position that participants developed either fully abstracted representations of the two analogs (a), or production of a decision rule for when to apply each solution strategy (b). Facilitory similarity problems were solved at a higher rate than misleading similarity problems in both conditions, which indicates that even the high cuing participants had stored problem contextual features as well as the mathematical structure. At the same time, the interaction between the effects of surface similarity and instructional condition seems to rule out the third possibility. Despite the effect of misleading surface similarity on the high cuing participants, the effect is relatively less than on the minimal cuing participants. If high cuing led to better retention for the test problems, as indicated in possibility three above, one might expect that the effect of surface similarity would be even greater for the high cuing condition. But, this is not the case. We do concede, however, that greater retention for the test problems could allow participants to more easily compare both surface and mathematical structure between instructed and posttest problems. Finally, these data do seem to support the fourth possibility, that the higher cuing teaches students to reason more like experts and improves their ability to analyze the structural properties of a new problem.

Experiment 2 extends these data in several ways, allowing for a better treatment of these interpretations in the final discussion. A new mathematical content area was used to explore the generalizability of Experiment 1 findings, particularly because permutation and combination problems have been used extensively in the analogy learning literature. Experiment 2 involved teaching a limit to the linearity assumption of proportional reasoning, which is another mathematical content area in which undergraduates are well known to demonstrate a lack of success, but is included on the GRE and so is at a suitable level for undergraduate students. Second, we sought to generalize across the two most common types of mathematical analogies in US instruction (Richland et al., 2004). Experiment 1 studied the effects of cuing on a comparison between two problem types, and Experiment 2 examined the effects of a comparison between two solution strategies. Finally, Experiment 2 included a delayed posttest in order to measure the impact of cuing over a more substantial delay. As in Experiment 1, posttests measured learners' ability to solve test problems that (a) appeared similar to instructed problems, and (b) in which the appearance and mathematical correspondences were cross-mapped.

6. Experiment 2

Experiment 2 compared undergraduates' learning from a highly versus minimally cued instructional analogy between two strategies for solving a proportion problem. This experiment also included a control condition in which participants were actively engaged in the instruction but there was not an analogy presented. We hypothesized that the highly cued instructional analogy would lead to the greatest ability to map instructed problems to test problems on the basis of structure.

6.1. Method

6.1.1. Participants

Participants were 80 undergraduates who participated for partial course credit (62 = females, $M = 21$ years). Four participants were excluded for failing to return on the second day of the study. Data for the remaining 76 participants were included in analyses.

6.2. Materials

6.2.1. Pretest

A pretest was included in the current study to determine the rate and pattern of participants' use of the linearity assumption. Because the experimental manipulation involved a comparison between two solution strategies, we determined that most participants would only solve the pretest problem once, and thus were less likely than in Experiment 1 to independently make their own comparison. Doing so could have impacted the affect of the instructional manipulation.

Pretest packets were constructed to measure both entry mathematics knowledge and demographic information. The packet contained a math problem that assessed learners' ability to design an appropriate proportion between two ratios: *Bob needs 6 h to paint a square wall with a side of 10 m. How many hours would he need to paint a square wall with a side of 5 m?* The problem also assessed the frequency of the linearity misconception: the likelihood of making an assumption of linearity between the two proportions (i.e., setting up the proportion $6/10 = x/5$). This is appropriate in some contexts, but in the case of an area, linearity does not hold. Rather, one must first calculate area and use that quantity in the proportion. The correct proportion should be $6/100 = x/25$.

6.2.2. Experimental videos

Videotaped instruction was developed to alert learners to the structural similarities and differences between making the linearity assumption and performing relevant calculations (area) before setting up a proportion. Three videos were designed. Two were experimental videos that provided an instructional comparison between these structurally related but different solution strategies. In both versions the teacher showed the use of linearity and then the correct solution strategy; however, these videos differed in their use of cues to promote analogical encoding. One version was a Low Cuing video, in which the linearity strategy was taught and recorded on the board, labeled incorrect for this problem, erased, and then the second strategy was taught and recorded on the board. The second video, High Cuing, taught the same two strategies to the same problem, but supplied several educationally realistic levels of cuing supplied to ensure learners' attention to the relational structure. The linearity strategy was taught and recorded on the board, but participants were not told whether it was correct or incorrect. It was left on the board while the correct strategy was demonstrated on the right side of the board. The two strategies were recorded on the board in a spatially aligned format (i.e., the problem was written in the same way and the proportions were written at the same spatial level).

In a procedure modeled after that introduced by Gick and Holyoak (1983) and that has been widely used since, participants were then given the opportunity to map correspondences between the two strategies and make a judgment about which strategy was correct. The video stopped and they were handed a worksheet with four questions. The first three prompted them to compare the two strategies on the basis of structure and map objects between the two proportions. The last question asked them to state which strategy they believed was accurate. When the experimenter determined that the participant was finished, the video resumed and the recorded instructor stated the right solution was correct, and used comparative gesture to highlight the relevant structural difference between the strategies (the calculation of area before setting up the proportion).

A third condition, active participation control, was designed to ensure that any performance differences between the participants in the two cuing conditions were not due only to subjects' active participation in the instruction. Since laboratory-based studies are plagued with participants' low motivation levels, and because active participation in instruction is known to facilitate learning, we developed a videotaped instruction that did not use a structural comparison, but did teach the correct solution strategy and engaged learners' in participating. In this videotape the instructor wrote the problem on the board and demonstrated how to set up the correct proportion. Since most participants used the linearity strategy on the pretest problem it was possible that this instruction invoked a mental comparison between these strategies, but such comparison was not supported by the instruction. After setting up the proportion, the video stopped and the experimenter handed participants a worksheet. As in the high cuing condition, the worksheet included four questions. These asked the participant questions about the instruction and asked them to perform the calculations necessary to solve the proportion. Crucially, these active participation prompts did not cue the learners to attend to any structural differences between the linearity and accurate solution strategies. After the experimenter determined that participants had finished, the video resumed and the recorded instructor performed the calculations and gave the correct answer. The details of the videotape construction are available in Appendix B.

6.2.3. Posttest

Two counterbalanced posttests were created so that each included two target problems alternating with three distractor problems. For each participant, one of the posttests was used as an immediate posttest, and the other was used as a week-delayed posttest. The posttests given at the two time periods were counterbalanced versions of the same test, so half of the participants got one version immediately and the second after a delay, and vice versa. Thus the same problems were distributed for all participants across the two time periods to ensure maximum comparability between the two tests.

As in Experiment 1, posttest questions orthogonally manipulated surface similarity and structural similarity. On each test, one posttest problem appeared similar to the instructed problem and could be solved using the taught solution strategy that avoids the linearity assumption (facilitory similarity). The second problem was set in the same surface context (painting) and thus appeared similar to the instructed problem, but the taught instruction was not relevant and the linearity assumption was actually accurate (misleading similarity). In fact this was an easier test problem, but we predicted that learners who were less effective at processing the structure of the target analog would over-extend the instruction to this case.

Scoring procedures were the same as in Experiment 1. Problem solutions were scored both for their accuracy and the type of error made. As in Experiment 1, calculation errors were not taken into account in the accuracy ratings. Second, errors were coded as other errors or as cross-mapping errors, since the latter were considered diagnostic of failures to appropriately discriminate between analogs on the basis of structure. Errors were coded as cross-mapping errors when participants made the linearity assumption when not appropriate, or applied the instructed strategy when linearity was appropriate. All participants who made such an error were given a 1, while those who scored correctly or made a different type of error were scored 0.

6.2.4. Procedure

In a between-subjects design, participants were randomly assigned to one of the three video conditions, High Cuing ($N = 26$), Minimal Cuing ($N = 23$), and active participation control ($N = 27$). An experimenter tested each participant individually at a computer. In all conditions, the experimenter administered the paper and pencil pretest, which was untimed. Afterwards, the experimenter removed the pretest and asked the participant to wear headphones that were provided. The experimenter then started the videotaped lesson on the computer. As noted above, in two of the conditions the experimenter administered a worksheet between two parts of the video. After the instructional video was complete, in all conditions participants were given one of two counterbalanced posttests. Testing was untimed. One week later, participants returned and completed the alternative posttest, again individually. The delayed posttest was also untimed.

6.3. Results

As in Experiment 1, accuracy data are described first, followed by data from cross-mapping errors.

6.3.1. Pretest

Pretest data are reported first to assess the base rates of performance across participants, and to verify that participants would over-extend the linearity assumption, as predicted by the larger body of research on undergraduates' mathematics knowledge. Additionally, the pretest allowed for ensuring random assignment of participants across the experimental conditions. Results from the pretest suggest that the majority of participants in all conditions did reveal this misconception. Twenty-three percent of participants ($n = 18$) overall correctly solved the problem on the pretest, whereas 64% made the expected linearity error ($n = 48$). The error rates were distributed across all three conditions. There were no differences between conditions on the pretest accuracy rates, $F(2, 73) = 1.1$, $p = .34$, $\eta_p^2 = .03$, nor of making the linearity assumption error, $F(2, 73) = 1.0$, $p = .36$, $\eta_p^2 = .028$. Participants were not excluded from further analyses if they gave a correct answer, since we found that their performance on the final and delayed posttests revealed the use of the misconception at comparable rates to those who used it on the pretests, and they were all similarly impacted by surface similarity.

6.3.2. Posttest accuracy

Posttest data were next analyzed to assess the impact of the experimental manipulation on retention of the instructed strategy (facilitory similarity problems) and extension to problems that were unlike the instructed problems (misleading similarity problems). Immediate and delayed posttest data were both included in an omnibus ANOVA to determine whether the performance patterns changed over time with forgetting. The independent variable, instructional condition, had three levels: high cuing, minimal cuing, active participation control. The model first tested whether the manipulation of surface similarity within the task materials was successful. As in Experiment 1, there was a main effect of problem type, $F(1, 73) = 69.7$, $p < .001$, $\eta_p^2 = .48$, meaning that varying the surface similarity of problems did impact the cognitive demands of the facilitory versus misleading problems. Accuracy was proportionally higher overall for facilitory similarity problems (immediate test: $M = .95$, SD = .28, delayed test: $M = .82$, SD = .39) than for misleading similarity problems (immediate test: $M = .32$, SD = .47, delayed test: $M = .46$, SD = .50). However, there was not a significant interaction between problem similarity type and condition, $F(2, 73) = 1.14$, $p = .\ 33$, $\eta_p^2 = .03$. Unlike in Experiment 1, condition effects were visible on both facilitory and misleading similarity problems (main effect result below), so these data are collapsed for the subsequent analyses of variations in learning by condition.

Interestingly, there were also few effects of the posttest delay. The main effect of delay of test was not reliable, $F(1, 73) = .04$, $p = .84$, $\eta_p^2 = .00$, as there was not a large overall difference between performance on the immediate and delayed posttest (see Fig. 4a versus b). There was also no interaction between test delay and instructional condition, $F(2, 73) = 1.14$, $p = .33$, $\eta_p^2 = .03$, indicating that any effects of the instructional manipulations are not easily interpreted as differences in retention or retrieval access. There was an interaction between problem similarity type and test delay $F(1, 73) = 11.7$, $p < .01$, $\eta_p^2 = .04$, but the pattern is reflective of the stimuli. The solution strategy that was correct for the instructed problem showed some forgetting (evident in the decrease over time in accuracy on facilitory similarity problems), whereas the linearity misconception that had been shown to be incorrect on the instructed problem underwent some reinstatement after a delay (evident in the small increase in performance on misleading similarity problems). We interpret this interaction as evidence that as participants increasingly forgot the instruction, they reinstated the misconception error many had revealed on the pretest. The three-way interaction between test delay, problem similarity type, and instructional condition was not reliable, $F(2, 73) = .59$, $p = .56$, $\eta_p^2 = .02$.

The omnibus model revealed a main effect of condition $F(2, 73) = 8.71$, $p < .001$, $\eta_p^2 = .19$. Mean scores for the high cuing condition were higher than either of the other instructional conditions for both problem types immediately and at a delay. These apparent differences between conditions were further explored using a planned contrast analysis. Since neither time nor problem similarity type was significantly related to instructional condition, the following pair-wise comparisons collapsed across those factors. These analyses revealed an overall difference between the conditions, $F(1, 73) = 8.71$, $p < .001$, $\eta_p^2 = .19$, with specific variations between the high and minimal cuing conditions (mean difference = .23, $p < .001$), and the high cuing and active participation control conditions (mean difference = .19, $p < .01$), but not between the minimal cuing and active participation control conditions (mean difference = .04, $p = .77$).

The high cuing condition thus led to higher performance than either the low cuing or control conditions, but there was very little difference between the low cuing and control conditions. This finding suggests that without high cuing during training, the comparison of two analogs was not more effective than an engaging lesson on one analog only. Further, these data suggest that directing novices' attention to relational structure during an instructional comparison between confusable analogs increased later accuracy and appropriate use of a relevant source analog.

Finally, paired *t*-tests were conducted only on the high cuing participants to determine whether, as in Experiment 1, there was

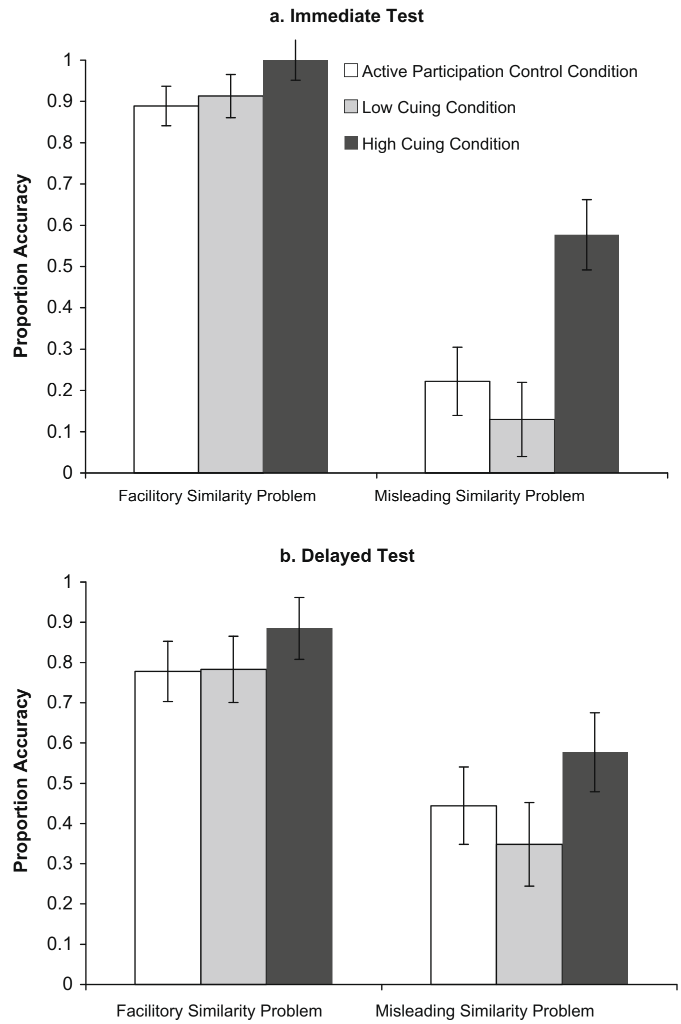


Fig. 4. Impact of support for active student participation in analogical instruction on posttest accuracy immediately and at a delay (Experiment 2).

a difference in cuing between facilitory and misleading similarity problems. This comparison provided an important test of whether these participants solved the target problem based on a purely schematic/conceptual understanding or by retrieving the source representation. The analyses revealed a difference both immediately, $t(25) = 4.2$, $p < .001$, $d = 1.7$, and at a delay $t(25) = 2.5$, $p < .05$, $d = 1.0$, such that scores were higher on facilitory similarity problems than on misleading similarity problems (immediate means = 1.00 versus .58, and at a delay, means = .88 and .58). This indicates that participants even in the high cuing condition were retrieving the instruction as a source analog while solving target problems.

6.3.3. Cross-mapping errors

An omnibus ANOVA was next conducted using cross-mapping errors as the dependent variable to examine whether the patterns observed in the accuracy data were explained by participants' failure to accurately map known analogs to new posttest problems. As with accuracy, all three conditions were first examined together to see the overall relationship between level of comparative instruction and performance on facilitory versus misleading similarity problems. A 2 (problem similarity type) $\times$ 2 (test delay) ANOVA was performed with instructional condition (three levels) as a between-subject factor. As was found for accuracy, there was a significant difference between problems, $F(1, 73) = 37.3$, $p < .001$, $\eta_p^2 = .34$, such that the rate of cross-mapping errors was lower on facilitory similarity problems overall (immediate test: $M = .05$, SD = .22, delayed test: $M = .16$, SD = .37) than for misleading similarity problems (immediate test: $M = .47$, SD = .50, delayed test: $M = .46$, SD = .52).

Again, there was not a main effect of test delay, $F(1, 73) = .04$, $p = .84$, $\eta_p^2 = .00$, since there was not a large overall difference between performance on the immediate and delayed posttest (see Table 2 for means). The interaction between problem similarity type and test delay was not reliable, $F(1, 73) = 1.8$, $p = .19$, $\eta_p^2 = .02$, nor was the three-way interaction between test delay, problem similarity type, and instructional condition, $F(2, 73) = .00$, $p = .99$, $\eta_p^2 = .00$.

In addition, there was no reliable interaction between time and condition, $F(2, 73) = .15$, $p = .86$, $\eta_p^2 = .00$, indicating that any effects of the instructional manipulations are not easily interpreted as differences in retention or retrieval access, since performance was fairly comparable immediately after training and after a week delay. Also, like in the accuracy data, there was not a significant interaction between problem similarity type and condition, $F(2, 73) = 1.0$, $p = .37$, $\eta_p^2 = .03$. However, the main effect of condition was reliable, $F(2, 73) = 4.89$, $p = .01$, $\eta_p^2 = .12$. As shown in Table 2, mean rates of cross-mapping errors for the high cuing condition were lower than for either of the other instructional conditions, for both problem types immediately and at a delay. These apparent differences between conditions were further explored using a planned contrast analysis. Since neither test delay nor problem similarity type were significantly related to instructional condition, the following pair-wise comparisons collapsed across those factors. These analyses revealed an overall difference between the conditions, $F(1, 73) = 4.89$, $p = .01$, $\eta_p^2 = .12$, with specific variations between the high and low cuing conditions (mean difference = .15, $p < .05$), and the high cuing and active participation control conditions (mean difference = .16, $p < .01$), but not between the low cuing and active participation control conditions (mean difference = .006, $p = .92$). The high cuing condition thus led to fewer cross-mapping errors than either the low cuing or control conditions, but there was very little difference between the low cuing and control conditions. This finding bolsters the pattern revealed in the accuracy data, suggesting that without high cuing during training, the low cued comparison during training was not more effective at promoting schematization than an engaging lesson using one analog only. Further, these data suggest that directing novices' attention to relational structure during an instructional comparison between analogs reduced future cross-mapping errors, and in particular, made learners less distracted by cross-mapped featural similarity.

6.4. Discussion

Experiment 2 conceptually replicated and extended findings from Experiment 1. Overall, results in Experiment 2 show that high supportive cues for an instructional comparison had direct implications for participants' future ability to discriminate between previously unseen potential analogs on the basis of structural correspondences rather than surface features. Importantly, the role of support was crucial to whether or not the instruction was effective at promoting this type of flexible problem solving. Participants in all conditions scored fairly well on problems that appeared similar to instructed problems immediately after instruction (over 90% accuracy) and there was not a significant effect of forgetting after a delay. Thus all instruction was fairly potent. However, the crucial finding was that performance on misleading similarity problems differed significantly between conditions depending on the level of instructional cuing provided during training. In the General Discussion we consider alternative explanations for this effect.

The mere inclusion of a comparison in instruction was not enough to promote expert-like reasoning, as evidenced by the results from participants in the low support condition. Participants in that condition saw the same correct instruction as did participants in the high support condition and were explicitly shown why the linearity assumption did not work in this case before being taught the correct strategy; yet when compared to the high support condition, these participants were significantly less well able to differentiate when to use each strategy. This gap was observed even immediately after training, and the differences remained after a substantial delay.

Moreover, active participation in the instruction was not enough to ensure participants' ability to identify and discriminate between analogs on the posttest. Participants in the control condition actively participated in the construction of the correct solution, but were less able to identify the key structural properties of potential analogs either immediately or at a delay than participants in the high cuing condition.

Altogether, these data indicate that dynamic, ecologically-valid visual and spatial cuing during training impacted learners' interactions with posttest problems encountered subsequent to instruction. Specifically, high cuing led to more successful discrimination between potential source solutions analogs. The impact of cuing at time of test is well known, but these data revealed that cuing during instruction directly impacted learners' later processing of target and retrieved source representations. The cues did not seem to impact retention or retrieval fluency for source analogs, as participants in all conditions were readily able to retrieve the relevant analog for solving the facilitory similarity problems, and performed higher on the facilitory than misleading similarity problems.

Table 2
Frequency of cross-mapped strategy use across problem type in Experiment 2.

Instructional condition	Facilitory similarity		Misleading similarity	
	Immediate test	Delayed test	Immediate test	Delayed test
High cuing condition	0 (0[a])	12 (33)	31 (47)	31 (47)
Low cuing condition	4 (21)	17 (39)	57 (51)	57 (59)
Active participation control	11 (32)	19 (40)	56 (51)	52 (51)

[a] Standard deviations listed in parentheses.

7. General discussion

7.1. Summary

The results of Experiments 1 and 2 both revealed that providing cues during an instructional comparison led to more flexible and schematized representations of taught concepts, as measured by increases in problem solving accuracy and decreases in cross-mapping errors, than did the same instruction with minimal cuing. This basic finding was demonstrated in two mathematical contexts in which the instructional analogy compared analogs that were confusable – representations that were structurally identical except for a crucial difference that caused one representation to require an additional transformation. Confusable analogs are related to the concept of a "near miss" – cases in which all features and structures are identical except for one crucial difference. However, the "near miss" concept has been used to emphasize crucial differences in object correspondences (see Ross & Kilbane, 1997; Nokes

& Ross, 2007), whereas in the present study the confusable analogs differed on a key structural relation.

Cuing during training impacted learning from comparisons between two contrasting problem analogs (Experiment 1), as well as comparisons between two alternative solutions to a single problem (Experiment 2). These results held both immediately and (tested in Experiment 2 only) after a week delay. Without such cues, the instructional analogy produced less flexible, expert-like reasoning, and was not more effective than a lesson in which participants were actively engaged in learning with one analog only (Experiment 2).

7.2. Possible mechanisms for flexible learning from analogies

The data from the present experiments provided insight into the possible mechanisms by which a highly cued instructional analogy could yield increases in participants' ability to differentiate between confusable analogs. Specifically, we sought to differentiate between four alternative characterizations of the impact of such cuing on novices' problem solving: (1) abstracted schematic representation of the two analogs, (2) production of a decision rule for when to apply each taught solution strategy, (3) improved retention of the individual problems, or (4) more expert-like processing of previously unseen problems (attention to structural versus surface features).

Data from the cross-mapping manipulation are especially informative. First, the main effect of surface similarity in both experiments eliminates the possibility that all, or even highly cued, instructional analogies automatically led to fully schematized knowledge representations of taught concepts. In both experiments, differences in accuracy across conditions were obtained between the facilitory and misleading similarity problems. If participants in the high cuing conditions were reasoning on the basis of purely schematic representations of the analogs, they would have shown comparable performance on the two types of problems. Instead, like participants in the other conditions, participants in the high cuing condition were misled by the irrelevant featural similarity. Thus, this evidence suggests that participants had at least partially stored the problem analogs intact, and were not using a fully abstract schematic representation.

It is possible that while stored source analogs were not fully schematized after high cuing, they were more schematized than following a more minimally cued instructional analogy. Perhaps if we had included prompts to evaluate participants' schema quality (as in Novick & Holyoak, 1991), we would have detected improved ability to represent the relational schemas in the high cuing conditions in spite of their use of the intact source analogs when mapping to target problems.

In addition, these same accuracy patterns demonstrate that participants did not appear to be using a discrimination rule to differentiate between the two types of problems. Vanderstoep and Seifert (1993) found that providing such a discrimination rule could help participants differentiate between permutation and combination problems; however, such a rule does not seem to be an inevitable consequence of comparative instruction, as we would have expected such a decision rule to be unaffected by irrelevant surface featural similarity.

The third possible explanation was that the instructional comparison led to better retention for the instructed problems, perhaps due to more enriched encoding. But as noted above, if retention were the differentiating factor one would expect a significant relationship between forgetting and condition, which was not the case. Furthermore, the relatively high performance on the facilitory similarity problems even at a week delay suggests that retention could not be the causal explanation. Participants in the low cuing and single analog control conditions did not show patterns of forgetting that were suggestive of minimally encoded, short-term retention for the problem solution analogs. It is possible, of course, that at longer delay differential effects of retention might be observed.

While the present data exclude schema-only, decision rule, or retention explanations, participants' performance in the high cuing conditions appears congruous with what one would expect from a more expert-like reasoner. Participants in the high cuing conditions may have not only gained more conceptual, structural knowledge representations of the instructed analogs, but also something more intangible – how to attend to relational structure when processing a new problem. The teacher used gesture, structural alignment, and visual cues, which in essence served to demonstrate the act of identifying and mapping structural correspondences between two analogs. Thus learners may have been trained in what it means to attend to relational structure in that particular mathematical context. If true, this interpretation generates two important predictions for future research. First, experts' ability to attend to relational correspondences may be distinguishable from domain knowledge within analogical thinking, since both high and low cuing conditions showed high retention of content and ability to solve facilitory similarity problems. Second, attention to relational structure may be trainable.

7.3. Broader theoretical implications

These results have theoretical implications for several bodies of educational and psychological literature on learning by analogy and classroom mathematics instruction. Drawing connections and comparing representations is core to mathematical thinking and generalizable learning (see National Mathematics Panel, 2008a, 2008b; Gallistel & Gelman, 2005; Hilbert, 1900; Polya, 1954; Skemp, 1976), but it is seriously underutilized in US classroom teaching (Hiebert et al., 2003; Richland et al., 2007). While the National Council of Teachers of Mathematics and disciplinary panels have long recommended using mathematical connections to deepen students' conceptual understanding, strategies for integrating these practices into teachers' normative routines have not been largely successful (Hiebert et al., 2003).

This paper used a novel approach to this problem by bridging cognitive science models of analogical reasoning (Gentner, 1983; Holyoak & Thagard, 1989/2002) with study of practices identified within everyday mathematics lessons taught in the United States and Internationally (Richland et al., 2007). These practices are feasible and require low resource investment by the teacher. Further, as evidenced by the experiments demonstrated here, benefits can be realized even without a greater time commitment. Thus the strategies identified here may provide a window into techniques that could help scaffold teachers' existing practices to become increasingly effective. Many studies reveal that simply invoking an instructional analogy, or another opportunity for comparing problem solutions, is not reliably effective (e.g., Gerjets et al., 2006; Gick & Holyoak, 1980, 1983; Rittle-Johnson & Star, 2007). Thus, the current tested strategies may be an important tool for ensuring that students benefit from the analogies as intended by the teacher.

Using a framework of analogical reasoning to consider mathematics instruction and the teaching of problem solving also provides insights into the distinction between surface and structural features of mathematics problems or representations. Much basic work in analogy indicates that children have great difficulty inhibiting responses to appearance and surface features that appear similar, in spite of being structurally irrelevant or misleading (e.g., Cho et al., 2007; Richland et al., 2006). This is crucial to mathematical thinking, which relies upon the ability to manipulate and engage with abstract concepts rather than focusing on surface appearance or problem context (Bransford et al., 2000; National Mathematics Advisory Panel, 2008a, 2008b). Thus drawing on insights from the literature on analogical reasoning allows for considering new strategies for drawing learners' attention to the key structural, mathematical correspondences in a learning context rather than surface features.

7.4. Implications for instructional practice

Finally, this interpretation has direct implications for US teaching practices. US teachers regularly use analogies in instruction but are less likely to use the high cuing techniques than their Asian peers (at least at the eighth-grade level). At the same time, they regularly cite their students' difficulty noticing mathematical commonalities between problems that appear different at a surface level (e.g., National Mathematics Advisory Panel, 2008a, 2008b). Findings from the current experiments suggest that modifying teachers' current practices of using analogy to include theoretically grounded strategies for cuing students' comparative thinking might at least partly address this problem. While the controlled laboratory context does not allow for fully generalizable interpretations of these results (they would need to be replicated in K-12 classrooms), the findings do provide important insights into strategies for optimizing teachers' current practices of analogy.

Of course, we do not minimize the challenges in incorporating such strategies into current practices. Any attempt to transform teaching is likely to fail unless teachers engage in deeply considered conceptual change, though the suggested modifications have several advantages. The identified cuing strategies align with experimental and theoretical principles but derive from classroom teaching practices in the US, Hong Kong, and Japan. Thus they are classroom feasible and would not require a full re-organization of currently normative teaching practices. They also would not require immense class-time commitments since US teachers already regularly invoke mathematical analogies. Even so, teachers would probably need to gain a nuanced understanding of their students' comparative thinking and the difficulties inherent in learning by analogy to be most effective.

We also caution against the interpretation that including these cuing strategies means simply using more direct instruction. Rather, we argue that these pedagogical tools free learners' resources and focus their attention onto key mathematical structure. By doing so, the teacher supports and facilitates the learners' own constructive, analogical thinking. This active sense-making process thus is made possible by the teacher's increased support, but it was the learners' own reasoning that impacted their more flexible representations and more expert-like treatment of new problems. As this is a key feature of children's struggles with mathematical proficiency (Kilpatrick, Swafford, & Findell, 2001), the present findings may have important implications for fostering mathematics learning.

Acknowledgments

We thank Keith Holyoak for helpful discussions and Yevgeniya Bukshpun for participation in study design and running participants. This material is partially based upon work supported by the Office of Naval Research under Grant N000140810186 and the National Science Foundation under Grant No. IIS-0757646. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Writing was supported by a National Academy of Education/Spencer Foundation Postdoctoral Fellowship to the first author.

Appendix A

	Minimal comparison condition	High comparison condition
Introduction	Begin both videos by introducing the concept of permutations. Do not mention combinations yet in either condition. Define permutations as: 'a math problem that allows you to find the number of possible arrangements of items in a particular order'	
Source analog	*"Suppose there are five people running in a race. The winner of the race will get a gold medal, the person who comes in second will get a silver medal, and the person who comes in third will get a bronze medal. How many different orders of gold, silver, and bronze can there be?"*	
Source solution	• Teacher asks how many runners could have come in first, and answers – 5. This number was written down first • Teacher explains that assuming someone has taken gold, how many runners could possibly come in second, and answers – 4 • Assuming the gold and silver are taken, the teacher questions how many runners could possibly have come in third, and answers – 3 • The three numbers are multiplied to reach the total number of possible orders	
Source solution representation (see Fig. 1a and b for screenshots of the source analogs on the videos)	• Three lines, with "gold," "silver," and "bronze" written underneath • Each number is filled out in turn – first gold (five possible), then silver (four possible) and bronze (three possible)	• Three lines, with "gold," "silver," and "bronze" written underneath. • Each number is filled out in turn – first gold (five possible), then silver (four possible) and bronze (three possible) • Once one slot is filled, a second row of three lines are drawn with the one filled in to leave a record of the thinking process. The last line has all three slots filled
Camera capture	Camera moves to capture only the free second half of the board, simulating the teacher erasing or changing an overhead slide to remove the source analog	Camera moves to capture the whole board, including the first and second analog

(*continued on next page*)

Appendix A (*continued*)

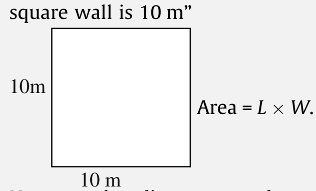
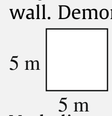
	Minimal comparison condition	High comparison condition
Transition	Teacher introduces combination problems, stating that he will explain them using the following scenario	Teacher introduces combination problems, stating that they are similar to permutations but have one main difference. He will explain them using the following scenario
Target analog	*A professor is choosing students to attend a special seminar. She has 11 students to choose from, but she only has four extra tickets available. How many different ways are there to make up the four students chosen to go to the seminar?*	
Target solution	• Teacher asks how many students could have received the first ticket, and answers – 11. This number was written down • Teacher explains that assuming someone has taken the first ticket, how many students could possibly take the second ticket, and answers – 10 • Teacher explains that using the same logic, only nine students could have taken the 3rd ticket • Finally, only eight students could have taken the 4th ticket • The four numbers are multiplied to reach the total number of possible orders • The total number of possible orders are divided by the	
Target solution representation (see Fig. 2 for screen shots of the source analogs in the videos)	• Four lines, with "ticket 1," "ticket 2," and "ticket 3" and "ticket 4" written underneath. • Each number is filled out in turn – 11, 10, 9, 8	• Four lines, with "ticket 1," "ticket 2," and "ticket 3" and "ticket 4" written underneath. • Each number is filled out in turn – 11, 10, 9, 8 • Once one slot is filled, a second row of three lines are drawn with the one filled in to leave a record of the thinking process. The last line has all three slots filled
Cues supporting comparison	Source and target analog presented in immediate sequence. The solution strategy and visual representation was the same for the two analogs save for the last target analog step	Source and target analog presented in immediate sequence. The solution strategy and visual representation was the same for the two analogs save for the last target analog step. Teacher referenced the source analog at each step of the target analog, using comparative gesture to the relevant element of the source analog. The teacher also noted when the final step in the target solution did not map to the source solution

Appendix B

	Comparison: high cuing condition	Comparison: low cuing condition	Active participation control condition
Introduction	Begin all videos with the same introduction, stating that this would be a problem-solving lesson, using one particular problem as an example		
Problem	Teacher writes on board and reads out loud: *Bob needs 6 h to paint a square wall with a side of 10 m. How many hours would he need to paint a square wall with a side of 5 m?*)		
Source analog solution strategy	Teacher describes an incorrect but common solution strategy: • Teacher states that one way to set up the proportion is to put hours in the numerator and meters in the denominator using the numbers given in the problem; Bob needs 6 h to paint a square wall with a side of 10 m. So we get: $\frac{6\ \text{h}}{10\ \text{m}}$ • Teacher restates the question – "*How many hours* will he need to paint a square wall that is *5 m?* $\frac{6\ \text{h}}{10\ \text{m}} = \frac{?\ \text{h}}{5\ \text{m}}$ • Teacher narrates and writes the cross-multiply strategy on the board, calculating the solution as 3 h. Teacher summarizes the reasoning behind this proportion: this makes sense because the side of a square wall of 5 m is half of the side of a square wall of 10 m"	–	
Comparative participation (Not active)		• Teacher states that this is not the correct way to solve the problem • Describes the issue – Bob is painting the *area* of the whole wall, not just the side	
Camera capture:	Source analog is left visible on the left side of the board	Board is cleaned and only the problem statement is visible	
Transition	"That's one way to approach the problem. Now lets look at this again		

Please cite this article in press as: Richland, L. E., & McDonough, I. M. Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology* (2009), doi:10.1016/j.cedpsych.2009.09.001

Appendix B (*continued*)

	Comparison: high cuing condition	Comparison: low cuing condition	Active participation control condition
Target analog solution strategy	• Teacher states that visualizing the problem should help students, and draws the following diagram: "Each side of the square wall is 10 m"  • Next to the diagram, teacher reminds them of the formula for area of a square, and writes the equation: 10 m × 10 m = 100 m^2 • Reminds them that they are trying to figure out how long it would take for Bob to paint a wall that has a *side* of 5 m, so they would need to now calculate again how much space he is painting. So we need to calculate the area of the second wall. Demonstrates this with a diagram:  • Verbalizes and writes the calculation for area of this square: 5 m × 5 m = 25 m^2 • Walks the students through setting up the ratio: hours in the numerator and meters squared in the denominator: $\frac{6\ \text{h}}{100\ \text{m}^2}$ • Teacher restates the main question with emphasis: How many hours will he need to paint a *25 m^2* wall? $\frac{6\ \text{h}}{100\ \text{m}^2} = \frac{?}{25\ \text{m}^2}$		
Non-comparative active participation			Active participation: teacher asks students to write and solve the problem themselves on their own piece of paper, showing their work. (video pauses)
Completion of target solution strategy	• Teacher narrates and writes the cross-multiply strategy on the board, calculating the solution as 1.5 h. • Summarizing, she states that they calculated using this ratio that Bob needs 1.5 h to paint a wall that's 25 m squared. Or in another words, it takes Bob 1.5 h to paint a wall with a side of 5 m • On board: 1.5 h to paint 5 m		
Comparative participation (active)	(*Video paused*) Participants given a sheet of paper with prompts to map correspondences between the two solution strategies, and asked to decide which one is correct (*video resumes*). Teacher states that the strategy on the left is not correct because Bob is painting the whole area, not the edge. Gestures between source and target strategies		
Conclusion	In strategy one, you have hours per meter, which is the time it takes to paint a side of a square wall. In strategy two, you have hours per meter squared, which is the time it takes to paint an area of the square wall. The first solution strategy is not realistic [Cross out]. We paint the whole wall (or in other words, the area) of the wall and not one side of a wall	So, we find that in order to calculate the correct number of hours, we need to first calculate the area of the wall, and then the time it takes to paint that area	
Cues supporting comparison	Two solutions presented in serial, both to the same problem, both using a ratio	Two solutions presented in serial, both to the same problem, both using a ratio. Solutions visible on the board simultaneously, teacher gestures between strategies when discussing their accuracy	–
	Now you'll be doing a set of problems on your own. Good luck and please show all your work		

References

Bassok, M., & Holyoak, K. J. (1993). Pragmatic knowledge and conceptual structure: Determinants of transfer between quantitative domains. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 68–98). Norwood, NJ: Ablex Publishing.

Blessing, S. B., & Ross, B. H. (1996). The effect of problem content on categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 792–810.

Bransford, J. D., Brown, A., & Cocking, R. (Eds.). (2000). *How people learn: Mind, brain, experience and school, expanded edition*. Washington, DC: National Academy Press.

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

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127*(4), 355–376.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1147–1156.

Chen, Z., & Klahr, D. (1999). All other things being equal: Children's acquisition of the control of variables strategy. *Child Development, 70*(5), 1098–1120.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Chi, M. T. H., & Ohlsson, S. (2005). Complex declarative learning. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 371–399). New York, NY: Cambridge University Press.

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*(6), 1445–1455.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching, 30*(10), 1241–1257.

Duit, R. (1991). On the role of analogies and metaphors in learning science. *Science Education, 75*, 649–672.

English, L. (Ed.). (2004). *Mathematical reasoning: Analogies, metaphors, and images*. Hillsdale, NJ: Erlbaum.

Gallistel, C. R., & Gelman, R. (2005). Mathematical cognition. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 559–588). New York, NY: Cambridge University Press.

Gentner, D. (1983). Structure-mapping – A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*, 393–408.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science, 10*, 277–300.

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations. *Learning and Instruction, 16*(2), 104–121.

Gick, M. L., & Holyoak, K. L. (1980). Analogical problem solving. *Cognitive Psychology, 15*, 306–355.

Gick, M. L., & Holyoak, K. L. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.

Goldstone, R. L., & Medin, D. L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 29–50.

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

Hilbert, D. (1900). *Mathematical problems.* Lecture delivered before the International Congress of Mathematicians, Paris, France.

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

Holyoak, K. J., & Thagard, P. (2002). Analogical mapping by constraint satisfaction. In T. A. Polk & C. M. Seifert (Eds.), *Cognitive modeling* (pp. 849–909). Cambridge, MA: MIT Press.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Kolodner, J. L. (1997). Educational implications of analogy: A view from case-based reasoning. *American Psychologist, 52*(1), 57–66.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

National Mathematics Advisory Panel (2008a). *Foundations for success: The final report of the national mathematics advisory panel*. Washington, DC: US Department of Education.

National Mathematics Advisory Panel, (2008b). *Reports of the task groups and subcommittees*. Washington, DC.

National Research Council (2001). Adding it up: Helping children learn mathematics. In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), *Mathematics learning study committee, center for education, division of behavioral and social sciences, and education*. Washington, DC: National Academies Press.

Nokes, T. J., & Ross, B. H. (2007). Facilitating conceptual learning through analogy and explanation. In L. Hsu, C. Henderson, & L. McCullough (Eds.), *2007 Physics education research conference* (pp. 7–10). American Institute of Physics Conference Proceedings.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 510–520.

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. G. W. C., & Vanmerriënboer, J. J. G. (1994). Instructional-control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*, 351–371.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., et al. (2007). *Organizing instruction and study to improve student learning (NCER 2007-2004)*. Washington, DC: National Center for Education Research, Institute of Education Sciences, US Department of Education. Retrieved <http://ncer.ed.gov>.

Perry, M., Vanderstoep, S. W., & Yu, S. L. (1993). Asking questions in first-grade mathematics classes: Potential influences on mathematical thought. *Journal of Educational Psychology, 85*, 40–41.

Polya, G. (1954). *Mathematics and plausible reasoning. I. Induction and analogy in mathematics. II. Patterns of plausible inference*. Princeton, NJ: Princeton University Press.

Reed, S. K. (1987). A structure–mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 124–139.

Reed, S. K. (1989). Constraints on the abstraction of solutions. *Journal of Educational Psychology, 81*, 532–540.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*, 15–22.

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., Holyoak, K. J., & Stigler, J. W. (2004). Analogy generation in eighth grade mathematics classrooms. *Cognition and Instruction, 22*, 37–60.

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*, 249–271.

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogy in the mathematics classroom. *Science, 316*, 1128–1129.

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.

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 629–639.

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.

Ross, B. H., & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 42–55.

Ross, B. H., & Kilbane, M. C. (1997). Effects of principle explanation and superficial similarity on analogical problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 23*, 427–440.

Scheiter, K., Gerjets, P., & Catrambone, R. (2006). Making the abstract concrete: Visualizing mathematical solution procedures. *Computers in Human Behavior, 22*, 9–25.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 5*, 484–494.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*, 475–522.

Skemp, R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*, 20–26.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load and selective attention as factors in the structuring of technical material. *Journal of Experimental Psychology: General, 119*, 176–192.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59–89.

Tohill, J. M., & Holyoak, K. J. (2000). The impact of anxiety on analogical reasoning. *Thinking & Reasoning, 6*, 27–40.

Vanderstoep, S. W., & Seifert, C. M. (1993). Knowing 'how' versus knowing 'when': Improving transfer of problem-solving principles. *Journal of the Learning Sciences, 3*, 93–111.

Viskontas, I. V., Morrison, R. G., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (2004). Relational integration, inhibition and analogical reasoning in older adults. *Psychology & Aging, 19*, 581–591.

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

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*, 137–166.

Please cite this article in press as: Richland, L. E., & McDonough, I. M. Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology* (2009), doi:10.1016/j.cedpsych.2009.09.001