Given the Choice, Students Select an Inferior Training Method For Algebra Problems

Robert S. Ryan
Blake A. Miller
Steven J. Craig
Elliot J. Fitzpatrick
Christopher M. O'Donnell

Kutztown University

Presented at the 21st Annual Convention of the Association For Psychological Science, San Francisco, CA, May 22 – May 25, 2009

Correspondence:

Robert S. Ryan
Box 730, Psychology Department
Kutztown University
Kutztown, Pennsylvania 19530
rryan@kutztown.edu

Given the Choice, Students Select an Inferior Training Method For Algebra Problems

Participants were pretested on algebra problems and then selected which of two training methods they believed would help them improve on a posttest. Next, they were trained by one of the two methods and took the posttest. Participants were more likely to select an inferior than a superior training method.

This paper addresses a potential difficulty that could interfere with a tactic that would otherwise be useful for helping people learn to solve algebra word problems. The tactic is comparing one training example to a second one that is analogically similar to it. Previous research has shown that learning from two examples that are both analogically similar to a test problem is better than learning from just one (Gick & Holyoak, 1983). When two examples are both analogically similar to the same test problem, then they are also analogically similar to one another. Gick and Holyoak found that comparing them, in addition to learning to solve them, makes people even more likely to later notice their analogical similarity to a subsequent test problem. Their research, however, used puzzle-type problems, not algebra word problems. Furthermore, a difficulty could arise if there is more than one training example that could potentially be used as a comparison example to another. If one such additional training example possessed the required analogical similarity, but another did not, then the potential difficulty is being able to recognize which training example would be useful for the comparison process.

This potential difficulty might not appear to be a problem in a typical classroom. A typical classroom procedure is for students to be taught with examples that are all similar to one another. They are then tested on new examples that are similar to the training examples. However, the potential difficulty might arise either if students were studying on their own, or if a teacher used a procedure of mixing dissimilar examples. It has been shown, that, even though such mixing of dissimilar examples impedes initial acquisition, it has benefits for later retention and transfer (Schmidt & Bjork, 1992). For that reason, using mixed examples is being increasingly recommended. Therefore, people's ability to recognize a training example that would serve as a useful one to compare to another needs to be investigated.

The current study presents a direct test of students' ability to recognize that a second training example that was analogically similar to a first training example would be most useful for the purpose of comparing the two examples in order to facilitate transfer to a test problem. The study included two phases. The first phase consisted of a choice task. This task tested the ability of students to select a type of training example that should be useful as a comparison example. The second phase was a training study that examined the actual usefulness of comparing with the different types of training examples from which the students chose.

Method

Participants

The participants were 120 college students who participated for credit in their General Psychology course at Kutztown University.

Materials and Procedure

The first step in the procedure was a pretest. That was followed by the choice task, the training, and, finally, the posttest. The choice task involved the same analogical relationships that existed between the training examples and the test problems. Therefore, in order to ease the explanation of the choice task, I will present the training examples and test problems first. Then I will explain the materials and procedure for the choice task. Finally, I will explain the training procedure.

Training examples and test problems. There were two training conditions. One was the two analogy training condition, and the other was the one analogy training condition. The training examples always included either one or two source analogues. The analogues could always be solved by weighted averaging of ratios. The test problems were the target analogues. They could be solved using the same procedure as the analogue training examples, but they had a different cover story. In both training conditions, the first training example was always a source analogue that had a cover story about discounts, such as the one below.

> **A man wanted to buy a suit that was priced at $300, but it was discounted 10%. He also wanted to buy a pair of shoes that were priced at $200 and discounted 20%. He wanted to know what his overall discount percentage would be on the entire purchase if he bought both items.**

In the two analogy training condition, the second training example was always a second source analogue. That is, it would be a good analogy for the first training example and for the test problems. It would always have a cover story involving distance, speed, and time, but could be solved by the same procedure as the first training example. Here is an example of such a second source analogue.

> **A ship captain sailed his ship from port A to port B in 3 hours sailing at 30 mph. Then he continued on to port C sailing for 15 hours at 42 mph. Later, he wanted to return by the same route and he wanted the trip to take the same length of time, but he wanted to travel at the same speed for the whole trip. He wanted to know how fast he should travel on the return trip.**

In the one analogy training condition, the second training example was always a mere appearance problem. That is, it always had the same cover story as the first training example, but its solution

procedure was different. Therefore, it was not a good analogy for the first training example, or for the test problems. Here is an example of a mere appearance training example.

A woman wanted to buy a digital camera that was priced at $450 but was discounted 20% due to a sale. However, the camera also had some minor scratches on it, so she convinced the sales person to give her an additional 10% off. She wanted to know how much money she would be saving with the two discounts.

The test problems tested transfer to a cover story about mixtures. The following are the two test problems that were used as the two forms of the test, one for the pretest, one for the posttest, counterbalanced across subjects:

A butcher has one batch of 120 lbs. of ground beef that is 5% fat. He also has another batch of 240 lbs. of ground beef that is 20% fat. If he mixes the two batches, what will be the percentage of fat in the resulting 360 lbs. of ground beef?

A grocery store sells rice that is a mixture of white rice and brown rice. They have 150 lbs. of mixed rice that is 60% brown rice. If they combine it with 100 lbs. of mixed rice that is 10% brown rice, then what is the resulting percentage of brown rice in the whole 250 lbs. of mixed rice?

Choice task materials and procedure. After the pretest, the participants participated in a choice task, in which they were randomly assigned to one of two conditions (see Appendixes A and B). In both conditions they were asked to imagine that they were going to participate in an experiment in which they would be trained to solve some algebra problems, and would then take a test. Thus, they were imagining the experiment that had already begun with their pretest, and would actually continue with their training and posttest.

The choice task was a match to sample task. As the sample, they were shown the first training example presented above (the discount training example). They were told that it would be useful as a worked example for learning to solve the test problems because the way it was solved was similar in some way to the way the test problems would be solved.

As the alternatives to choose from to match to the sample, they were shown two other examples. One of the two choices was always completely dissimilar from the sample. The other choice was always either one or the other of the two potential second training examples (the analogy or the mere appearance example) described above. Which example was the second choice depended on their choice task condition.

In both choice task conditions, one of the second potential training examples, which we called an anomaly, was dissimilar from the first training example both in terms of surface features as well as in terms of its solution procedure. In fact, although it involved ratios, it did not involve weighted averaging of them. The following is the anomaly:

A very long train was traveling at 10mph. A man jumped on the back of the train as it passed. He walked across the tops of the cars towards the front of the train at 3mph as it moved down the track. After 2 hours he jumped off. He wanted to know how far he had traveled.

The other potential second training example was similar to the first training example in one way, but dissimilar in another way. Which way the example was similar depended on the condition to which the participant had been randomly assigned.

The first condition was called the two analogy condition because, if the participant chose as the potential second training example the one that was similar to the first, rather than the anomaly, then they would have indicated that they believed that two examples, both of which were analogies to the test problems, would be the most useful to compare. In the two analogy condition, the other potential training example was the distance, speed, and time training example presented above to illustrate the relationship between the test problems and the training examples (The problem that began, "A ship captain sailed his ship from port A to port B in 3 hours sailing at 30 mph. ..."). Thus, in the two analogy condition, the participant had to chose between a completely dissimilar example (the anomaly) and a good analogy.

The second condition was called the one analogy condition because, if the participant chose as the potential second training example the one that was similar to the first, then they would have indicated that they believed that it would be most useful to compare two examples that were only superficially similar (although one of those examples would still be a good analogy to the test problems). In the one analogy condition, the other potential training example was the mere appearance example that was presented above to illustrate the relationship between the test problems and the training examples (The problem that began, "A woman wanted to buy a digital camera that was priced at $450 ... "). It had the same surface features as the first training example, but the same solution procedure as the anomaly. In other words, its similarity to the first training example was merely one of appearance. It bore a surface similarity to the first training example, but not structural similarity. Thus, in the one analogy condition, the participant had to choose between a completely dissimilar example (the anomaly) and a mere appearance example.

The two choice examples were described as being potentially useful as a second training example. That is, the participants were told that one of those other two examples would be a useful example to compare to the first example because, like the first example, the way it was solved was similar in some way to the way the test problems would be solved. They were told that doing such a comparison would help them to notice the similarities in the solution procedure, which would then help them to use that solution procedure on the test problems. Of course, what they were being told was expected to be true only if they were in the two analogy choice condition. It was actually expected to be false if they were in the one analogy choice condition. The participants' task was to select which of the two potential second training examples would be the most useful one to use, specifically as a comparison example, to the first training example.

Great care was taken to insure that the participant understood what they were to use as the basis for their choice. Participants were run either singly, or only two at a time. After receiving the instructions for the choice task, the participant (or one of the two participants) was asked to explain the instructions in their own words. Often, the participant said that the basis for the

choice was to be which of the two potential second training examples was most similar to the first. The participant was told that that was not quite correct. The basis for the choice was to be which example was similar specifically in the way that it would be solved. Thus, it would be the most useful example to use as a comparison example to the first one. Then the participant (or the other of the two) was asked to explain the basis for the choice one more time. The experiment did not continue until the participants clearly understood the correct basis for the choice.

If the participants in the two analogy condition could recognize that the way that the similar example could be solved was, like the first training example, also "similar in some way to the way the test problems would be solved", then they should be likely to select it over the anomaly example more often than would be expected by chance (50% of the time). However, the "way" that it was similar was in the important way that it involved a weighted averaging procedure. If they could not recognize that important similarity, then they would be just guessing, and therefore have about a 50% likelihood of selecting it.

If the participants in the one analogy condition condition could recognize that the way that the mere appearance example could be solved was, like the anomaly example, different from the way the first training example would be solved, then they would be no more likely to select it than the anomaly example than would be expected by chance. Thus, they should select it only about 50% of the time. However, if they did not recognize that it was not similar to the first training problem in the important way of being a weighted averaging problem, then they might think that its having the same surface features as the first training example would make it more useful as a training example than the anomaly example. Therefore, they might select it more often than would be expected by chance.

It is important to note that if the participants could successfully recognize whether or not the second potential training example, that is, the analogy or the mere appearance example, could be solved by the same procedure as the first training example, then they could also notice how many useful training examples the training would provide. In the two analogy condition, given that the participant selected the similar problem, they would be selecting a method of training in which both training examples would be solvable in the same way as the test problems. In the one analogy condition, regardless of which second potential training example the participant selected, they would be selecting a method of training in which only one of the training examples would be solvable in the same way as the test problems. Therefore, being able to recognize whether or not their choices for a second training example could be solved in the same way as the first, would make it even more likely that a higher percentage of participants in the two analogies condition would select the analogy problem than the percentage of participants in the one analogy condition selecting the mere appearance problem.

Training procedure. The preceding analysis assumes that training with two analogies to the test problems would in fact be more useful than training with one analogy and one mere appearance problem. Previous research on analogies suggests that it would be (e.g., Gick & Holyoak, 1983). However, in order to provide a direct test, in the next phase of the experiment, half of the participants from each of the choice task conditions was randomly assigned to be trained in the two analogies training condition, and the other half was assigned to be trained in the one analogy training condition. The training consisted of worked examples, a guided practice

phase, and an unguided practice phase. Each phase used two or more pairs of examples. Each time the participants were trained with a pair of examples, the training included both solving the problems and comparing the two members of the pair. There were actually two different methods of comparison. However, because the methods of comparison are not the focus of this paper, they will not be discussed further here. The training was followed by the post test.

In the worked examples phase, the participants were shown two pairs of worked examples which were explained to them. The worked examples phase was not timed. In the guided practice phase they were given two more pairs of examples. The participants solved these examples while an experimenter insured that they got both examples correct, and helped them to arrive at the correct answer if needed. The guided practice phase was also not timed. In the unguided practice phase, the participants solved as many more pairs of examples as they could in 15 minutes.

The relationship between the members of the pairs of training examples paralleled the relationship between the problems in the choice task. In the two analogies training condition, both members of every pair were good analogies for the test problems. The two examples that were used as the analogically similar examples in the two analogy condition of the choice task were actually used among the worked examples. Other pairs of training examples, although not identical to the choice task problems, were of the same type. That is, they were always structurally similar to each other, although they were dissimilar in terms of surface features.

Similarly, in the one analogy training condition, one member of every pair was a good analogy for the test problems, and the other member was a mere appearance problem. The two examples that were used as the merely apparently similar examples in the one analogy condition of the choice task were actually used among the worked examples, and other similar pairs were used as further training examples. That is, they were always similar in terms of surface features, but they were structurally dissimilar.

After the training, the participants took the posttest. The procedure for the posttest was the same as for the pretest.

Summary. To summarize, the participants took a pretest on an algebra word problem requiring the use of weighted averaging of ratios. Then they were given a choice task to assess whether they would recognize the value of analogical similarity over mere surface similarity for purposes of comparing during training. Next, they were actually trained with pairs of examples that were either analogically similar or only similar in their mere appearance. Finally, they took a posttest to verify the superiority of the two analogy training.

Results and Discussion

Pretest Performance

Performance on the pretest was examined, as shown in Table 1, to determine if there were any significant differences in ability to solve the pretest problems between the participants who were later randomly assigned to the two training conditions. This difference was examined separately for the two choice task conditions to which the participants were later randomly assigned. All *p* values are for two sided Fisher exact tests for the difference between the percentage of participants solving the pretest problem in the two training conditions.

Table 1 shows that, using the .05 alpha level, there were no significant differences in ability between participants who were later randomly assigned to the two different training conditions regardless of the choice task condition to which they were later randomly assigned. However, for those subjects who were later assigned to the two analogies choice task condition, those who were later assigned to the one analogy training condition performed at least marginally better than those later assigned to the two analogies training condition. The possible implications of this difference will be addressed in the discussion of the posttest results.

Table 1
Percent Correct On The Pretest As a Function of Training Condition for Each Choice Task Condition (number of participants correct out of number of participants in each condition is shown in parentheses).

Choice Task Condition	Training Condition		
	Two Analogies	One Analogy	*p*
Two Analogies	10 (3/30)	31 (9/29)	.057
One Analogy	21 (6/29)	19 (6/32)	1.000
Total	15 (9/59)	25 (15/61)	.256

Choice task

In Table 2, choice task performance is shown as the percentage of participants who chose the similar problem, as opposed to the anomaly, as the best potential second training example to compare to the first training example. Thus, in the two analogy choice condition the table shows the percentage who chose the analogy example, and in the one analogy choice condition it shows the percentage who chose the mere appearance example.

The percentages are shown separately for those participants who were incorrect and correct on the pretest problem. All *p* values are for two sided Fisher exact tests for the difference in the percentage of participants choosing the similar problem in the two choice task conditions.

Table 2 shows that, overall, participants were more likely to think that a mere appearance problem was the best problem to compare to a good analogy than to think that a second good analogy was the best comparison problem. However, that pattern was less strong among the minority of participants who were correct on the pretest problem.

Table 2

Percent of Participants Who Chose The Similar Problem as a Function of Choice Task Condition For Participants Who Were Incorrect and Correct on Each Pretest Problem (number of participants who chose the similar problem out of number of participants is shown in parentheses).

Performance On the Pretest Problem	Choice Task Condition		
	Two Analogies *(Percent Who Chose the Analogy)*	One Analogy *(Percent Who Chose the Mere Appearance)*	*p*
Incorrect	53 (25/47)	82 (40/49)	.004 **
Correct	75 (9/12)	83 (10/12)	1.000
Total	58 (34/59)	82 (50/61)	.005 **

Note: ** indicates statistically significant at the $p < .01$ level.

Posttest Performance

Table 3 shows performance on the posttest problems as a function of training condition. The comparison of performance in the two training conditions is shown separately for the two choice task conditions. It is further broken down according to the participants' responses on the choice task. All *p* values are for two sided Fisher exact tests for the difference between the percentage of participants solving the posttest problem in the two training conditions.

Table 3
Percent Correct On The Posttest Problem As a Function of Training Condition for Each Choice Task Condition And For Each Possible Response on The Choice Task (number of participants correct out of number of participants in each condition is shown in parentheses).

Choice Task Condition	Training Condition: Two Analogies	Training Condition: One Analogy	*p*
Two Analogies			
Chose Anomaly	67 (10/15)	50 (5/10)	.442
Chose Analogy	80 (12/15)	74 (14/19)	1.000
Total	73 (22/30)	65 (19/29)	.580
One Analogy			
Chose Anomaly	83 (5/6)	80 (4/5)	1.000
Chose MA*	91 (21/23)	52 (14/27)	.004 **
Total	90 (26/29)	56 (18/32)	.004 **
Total of Choice Task Conditions			
Two Analogies	69 (41/59)		
One Analogy	72 (44/61)		
Total	81 (48/59)	61 (37/61)	.016 *

Note: MA stands for Mere Appearance.
Note: * indicates statistically significant at the $p < .05$ level.
Note: ** indicates statistically significant at the $p < .01$ level.

Table 3 shows that, overall, the participants in the two analogies training condition performed better on the posttest problem than those in the one analogy condition. The Fisher exact tests for the difference between the training conditions show that the overall difference was due to the very large difference (39 percentage points) for those participants that had been in the one analogy choice condition and who chose the mere appearance example. That was the largest difference for any of the subgroups.

In contrast, among those participants that had been in the two analogy choice condition and who had chosen the analogy example, the difference was only 6 percentage points. Among those participant, those who were trained in the one analogy condition outperformed their counterparts who chose the anomaly by 24 percentage points. One the one hand, those who were trained in the *two* analogies condition and who chose the analogy did not outperform those who chose the anomaly by a similarly large margin. Nevertheless, it has been shown that comparing unworked examples does result in abstracting a schema for the examples' structures (Kurtz & Lowenstein, 2007). Therefore, it is tempting to speculate that those participants who had been in the two analogies choice task condition and chose the analogy may have learned from comparing problems during their choice task, and thus performed better on the posttest than they otherwise would have, even if they were trained with only one analogy.

Table 3 also shows that, overall, the participants in the one analogy choice condition performed only slightly, and not significantly, better than those in the two analogies choice condition. A two sided Fisher exact test for the difference between those conditions yielded $p > .05$.

Because the advantage of the two analogy training was very evident for those participants who had been in the one analogy choice condition, but less so for those who had been in the two analogy choice condition, it appears that there might be an interaction between the training conditions and the choice task conditions. However, it is possible that this seeming interaction is an artifact, or at least partially an artifact, of prior differences in math ability as suggested by the pretest scores.

Very few participants were correct on the pretest and there were no differences in performance between the training conditions that reached the .05 criterion for statistical significance. However, the largest difference in pretest performance was between the two training conditions for the one analogy choice condition. This difference approached significance in a two sided Fisher exact test ($p = .057$). Therefore, it is possible that the pretest performance might account for, or at least partially account for, the seeming interaction. If so, then analyzing difference scores, rather than posttest scores, might eliminate the seeming interaction.

Because the difference scores could range from -1 to +1, it was possible to analyze them with an ANOVA. Therefore, in order to examine whether the choice task condition really interacted with the training task condition, and to confirm the main effect of the training conditions, after accounting for the pretest scores, I conducted a two way ANOVA on the difference scores, using choice task condition and training condition as factors.

Table 4 shows the mean difference scores as a function of training condition and choice task condition.

Table 4
Mean Difference Scores As a Function of Training Condition And Choice Task Condition.

Choice Task Condition	Training Condition		
	Two Analogies	One Analogy	Total
Two Analogies	.63	.34	.49
One Analogy	.69	.38	.52
Total	.66	.36	.51

The ANOVA showed that the two analogies training condition performed significantly better than the one analogy training condition, F (1, 116) = 7.67, p = .007, MSE = .355. There was no main effect of choice task condition and no interaction. Therefore, what appears to be a possible interaction in which the beneficial effect of the two analogies training condition occurred only for the one analogy choice task condition could be an artifact, or at least partially an artifact, of differences in pretest performance.

It is important to note, however, that the random assignment of participants to the choice task and training conditions was not blocked on whether they were correct or incorrect on the pretest. Therefore, the result described above does not rule out the possibility that learning from the choice task could also have had some influence on the posttest results.

General Discussion

The training phase of this study shows that, consistent with previous research using other types of problems (e.g., Gick & Holyoak, 1983), training that involves comparing two algebra word problems that are analogically similar to one another improves subsequent performance on an analogically similar test problem. However, the choice task phase of the study shows that, given the choice, people are more likely to believe that comparing two problems that merely appear similar to one another will be useful for improving the subsequent test performance.

This result suggests that if future instructional methods begin to increasingly employ mixed examples for training, then it is important to provide direct instruction in which examples should be analogically compared to benefit transfer, rather than leaving it up to the learner to attempt to make that determination.

Finally, another study should be conducted in which performance on the pretest is used as a blocking factor before randomly assigning participants to the choice task, and performance on the choice task should be used as a blocking factor before assigning participants to training

conditions. Doing so could show whether or not being assigned to a particular choice task condition also affects transfer independent of the training condition and prior ability. It could also show whether performance on the choice task is related to transfer. If the choice task conditions differentially affect transfer, then such a finding might suggest possible methods for the direct instruction that this study shows is important.

References

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

Kurtz, K. J., & Lowenstein, J. (2007). Converging on a new role for analogy in problem solving and retrieval: When two problems are better than one. *Memory and Cognition, 35,* 334-341.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207-217.

Appendix A

The Two Analogy Choice Task Condition

Please read the following problem.

A man wanted to buy a suit that was priced at $300, but it was discounted 10%. He also wanted to buy a pair of shoes that were priced at $200 and discounted 20%. He wanted to know what his overall discount percentage would be on the entire purchase if he bought both items.

The problem above is called "Problem A". You do not have to solve it. Instead, pretend that you are in an experiment in which you have to solve some test problems (like the nine problems you just tried). Pretend that "Problem A" was given to you as a *worked out* example because the way it was solved was similar in some way to the way the test problems would be solved.

Now imagine that you were given a chance to see *another* worked out example, Problem B. The way that Problem B was solved was *also* similar in some way to the way the test problems would be solved. Therefore, Problem A and Problem B could be similar to *one another*. Furthermore, in order to be able to *use* the similarities between Problems A and B to help you solve the test problems, it might be *useful* to compare Problems A and B.

If you were in that type of experiment, and Problem A above was given to you as one worked example, which of the two problems below would you think would be the most *useful* problem to be the *other* worked example to compare to Problem A? Please circle the problem you choose.

A very long train was traveling at 10mph. A man jumped on the back of the train as it passed. He walked across the tops of the cars towards the front of the train at 3mph as it moved down the track. After 2 hours he jumped off. He wanted to know how far he had traveled.

A ship captain sailed his ship from port A to port B in 3 hours sailing at 30 mph. Then he continued on to port C sailing for 15 hours at 42 mph. Later, he wanted to return by the same route and he wanted the trip to take the same length of time, but he wanted to travel at the same speed for the whole trip. He wanted to know how fast he should travel on the return trip.

Appendix B

The One Analogy Choice Task Condition

Please read the following problem.

A man wanted to buy a suit that was priced at $300, but it was discounted 10%. He also wanted to buy a pair of shoes that were priced at $200 and discounted 20%. He wanted to know what his overall discount percentage would be on the entire purchase if he bought both items.

The problem above is called "Problem A". You do not have to solve it. Instead, pretend that you are in an experiment in which you have to solve some test problems (like the nine problems you just tried). Pretend that "Problem A" was given to you as a *worked out* example because the way it was solved was similar in some way to the way the test problems would be solved.

Now imagine that you were given a chance to see *another* worked out example, Problem B. The way that Problem B was solved was *also* similar in some way to the way the test problems would be solved. Therefore, Problem A and Problem B could be similar to *one another*. Furthermore, in order to be able to *use* the similarities between Problems A and B to help you solve the test problems, it might be *useful* to compare Problems A and B.

If you were in that type of experiment, and Problem A above was given to you as one worked example, which of the two problems below would you think would be the most *useful* problem to be the *other* worked example to compare to Problem A? Please circle the problem you choose.

A very long train was traveling at 10mph. A man jumped on the back of the train as it passed. He walked across the tops of the cars towards the front of the train at 3mph as it moved down the track. After 2 hours he jumped off. He wanted to know how far he had traveled.

A woman wanted to buy a digital camera that was priced at $450 but was discounted 20% due to a sale. However, the camera also had some minor scratches on it, so she convinced the sales person to give her an additional 10% off. She wanted to know how much money she would be saving with the two discounts.