Feature Matching Facilitates Analogy Based Transfer

Robert S. Ryan
University of Pittsburg

Sean Snoddy
Kutztown University

When people compare examples of unfamiliar problems they often infer the underlying structure, which then facilitates subsequent transfer. However, when algebra word problems are compared, such an inference is more difficult because it could rely on recognizing analogically similar structures either in terms of a general principle or in terms of a solution procedure. In order to examine a possible method to overcome this difficulty, in the present experiment, participants compared algebra word problems either by matching features that corresponded in terms of a general principle, or by verbally explaining the steps of the solution procedure. The feature matchers outperformed the explainers at subsequent transfer to an especially difficult problem that required them to use an equation that they had not been taught. They did so even though their training task involved seemingly mindless pattern matching, whereas the explainers' task involved verbal explanation. Therefore, training exercises designed to foster transfer of algebra problem solving ability might benefit by using a feature matching approach, although making such a task meaningful to students might be a challenge.

KEYWORDS: Feature matching, Self –explanation, Problem comparison, Problem solving, Algebra word problems, General principles at test

Using an analogy to a previously solved problem is often a good way to solve an unfamiliar problem. Although people can recognize analogies in some situations, such as analogies between stories (Gentner, Rattermann, & Forbus, 1993), they have difficulty doing so with algebra word problems (Reed, Dempster, & Ettinger, 1985). One possible reason for the difficulty could be because analogies rely on recognizing the similarity of the underlying structures of two problems, and word problems could represent more than one kind of underlying structure.

Consider these two problems. Problem A: 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? Problem B: Two airplanes leave from the same city at the same time heading for the same destination. The first airplane flies for 2 hours at 150 mph. Then it encounters engine trouble and slows down to 100 mph. It flies for 3 more hours at 100 mph. The second airplane arrived at the destination at the same time as the first plane, but it flew at the same speed for the full 5 hours. How fast was the second airplane flying?

Both problems could be described as involving two initial amounts, each associated with a ratio, that could be combined to form a final amount, also with an associated ratio. Therefore, both problems have the principle of weighted averaging of ratios as an underlying structure. In both problems, the unknown is the final ratio. Therefore, the solution procedures of the two problems involve the same set of steps, and the solution procedure is another kind of underlying structure. In this case, as shown in Table 1, the problems are analogically similar at two levels. First, they form an analogy at the level of the principle of weighted averaging. That is, when you set up both problems as equations for weighted averaging, you see that the relationships between the corresponding elements (the arithmetic operations) are the same. Second, they form an analogy at the level of the solution procedures. That is, when you set up both problems as a set of solution steps, you see that the same thing happens.

If people can recognize the similarity between solution procedures, then they could use Problem A as an analogy to solve Problem B. But suppose Problem B had had one of the two initial ratios as the unknown. Then, as shown in Table 2, the problems would still form an analogy at the level of the principle of weighted averaging, but not at the level of their solution procedures. Therefore, in order to use Problem A to solve Problem B, one would have to recognize the similarity at the level of the principle of weighted averaging, and then use that principle to derive a new solution procedure to find the initial ratio. How, then, can examples like the ones above be used to train people to solve these kinds of problems in such a way that they recognize the underlying principle?

Table 1. *Two Problems That Are Analogically Similar Both In Terms of The Principle of Weighted Averaging And In Terms of Their Solution Procedures*

Principle of Weighted Averaging						
	First Initial		**Second Initial**		**Final**	
A.	150(.60)	+	100(.10)	=	(150 + 100)X	
	First Initial		**Second Initial**		**Final**	
B.	2(150)	+	3(100)	=	(2 + 3)X	
Solution Procedures						
Step 1						
A.	150 lbs.	*	.60	=	90 lbs.	**First Initial**
B.	2 hrs.	*	150 mph	=	300 mi.	**First Initial**
A.	100 lbs.	*	.10	=	10 lbs.	**Second Initial**
B.	3 hrs.	*	100 mph	=	300 mi.	**Second Initial**
Step 2	**First Initial**		**Second Initial**		**Intermediate**	
A.	90 lbs.	+	10 lbs.	=	100 lbs.	
	First Initial		**Second Initial**		**Intermediate**	
B.	300 mi.	+	300 mi.	=	600 mi.	
Step 3	**Intermediate**		**Final**		**Final**	
A.	100 lbs.	/	250 lbs	=	X	
	Intermediate		**Final**		**Final**	
B.	600 mi.	/	5 hrs.	=	X	

Table 2. *Two Problems That Are Analogically Similar In Terms of The Principle of Weighted Averaging But Not In Terms of Their Solution Procedures*

Principle of Weighted Averaging						
	First Initial		**Second Initial**		**Final**	
A.	150(.60)	+	100(.10)	=	(150 + 100)X	
	First Initial		**Second Initial**		**Final**	
B.	2(150)	+	3(X)	=	(2 + 3)120	
Solution Procedures						
Step 1						
A.	150 lbs.	*.	60	=	90 lbs.	**First Initial**
B.	2 hrs.	*	150 mph	=	300 mi.	**First Initial**
A.	100 lbs.	*.	10	=	10 lbs.	**Second Initial**
B.	5 hrs.	*	120 mph	=	600 mi.	**Final**
Step 2	**First Initial**		**Second Initial**		**Intermediate**	
A.	90 lbs.	+	10 lbs.	=	100 lbs.	
	Final		**First Initial**		**Intermediate**	
B.	600 mi.	-	300 mi.	=	300 mi.	
Step 3	**Intermediate**		**Final**		**Final**	
A.	100 lbs.	/	250 lbs	=	X	
	Intermediate		**Second Initial**		**Second Initial**	
B.	300 mi.	/	3 hrs.	=	X	

One way to examine this question would be to train people with different methods, using problems like those above as training examples, but without teaching them the principle of weighted averaging, and then test them on transfer problems that required using not just any new solution procedure based on the principle, but rather, a new procedure that required actually using the equation for weighted averaging. For example, imagine that after subjects were trained with problems like the ones above, they were presented with this transfer problem: A wine company has 40 gallons of wine that is 25% alcohol. They need to combine it with some wine that is 4% alcohol, so that the resulting wine will be 10% alcohol. How much of the 4% alcohol wine should be added to the original 40 gallons of wine? Although the training problems could be solved either by using an equation for weighted averaging or by just following a step-by-step procedure, this transfer problem virtually requires the use of this weighted averaging equation: 40(.25) + X(.04) = (40 + X).10. This is an especially difficult problem for most students. Pilot studies showed that almost no college students could solve this problem without training.

Therefore, if one method of training results in greater success on such transfer problems, then that shows that that training method supports people's ability to infer the principle. Ryan (2005) conducted such an experiment and showed that comparing problems by making a similarity judgment (i.e., are the unknowns the same in the two training problems, or different?) resulted in just such transfer as described above. However, before Ryan's subjects could explain which problem element was the unknown in each problem, they had to compare the problems and find the corresponding initial and final elements in both problems. Thus, his task involved both a verbal explanation component and a matching features component.

Explaining about what one is learning during training, called self-explanation, has been shown to be beneficial for learning and transfer (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, DeLeeuw, Chiu, & LaVancher, 1994; Pirolli & Bielaczyc, 1989). Therefore, one possible explanation for why Ryan's (2005) subjects succeeded at transfer could have been because the explaining that they did helped them to realize that different unknowns lead to different solution procedures, but that all of them are based on the same principle. On the other hand, determining which problem elements in both problems were the initial and final elements, in effect, maps the elements together as one does when forming an analogy (Clement & Gentner, 1991; Gentner, 1983), and in this case, the analogy would be an analogy between the underlying structure based on the principle of weighted averaging as shown in Tables 1 and 2. If this matching component helped the subjects to understand the weighted averaging principle on which the underlying structure of the problems was based, then that might have been what led to the transfer. In the present experiment, we disentangled these two possible explanations by using one training method involving verbal explanation, but no matching, and another involving matching, but no verbal explanation.

Method

Participants

The participants were 157 psychology undergraduate students at a major Midwestern university who participated as part of the requirements of their Introductory Psychology course. The data from 7 participants were not included because of procedural errors, leaving 150 participants' data in the analysis.

We collected the individual difference variables of gender, age, amount of time since the participants took their most recent algebra course, math SAT score, score on a 100 item vocabulary test (as a proxy for I.Q.), verbal SAT score, and college GPA.

The participants were 48.7% female. Their age ranged from 18 through 49, with an average of 21.09 (SD = 4.57) and a median of 20. The amount of time since the participants took their most recent algebra course ranged from currently enrolled to 29 years ago, with an average of 4.61 (SD = 3.82) and a median of 4. The math SAT scores ranged from 260 to 730, with an average of 556.29 (SD = 83.60), and a median of 560.

There were no differences across conditions for gender, X2 (2, N = 150) = 1.33, $p > .05$, age, $F(2, 147) < 1$, amount of time since they took their most recent algebra course, $F(2, 132) < 1$, or math SAT scores, $F(2, 94) = 2.03$, $p = .137$. Although not all participants provided data for all of those variables, the number of subjects for whom we had data did not differ across conditions by more than 5, and most differences were only 1 or 2.

However, as shown in Table 3, the match features participants had higher average vocabulary test scores, verbal SATs, and GPAs than the other two conditions. The differences between conditions were significant for all three variables, $F(2, 146) = 3.40$, MSE = 372.26, $p = .036$, F (2. 94) = 5.73, MSE = 8799.21, $p = .005$, and F (2, 100) = , MSE = .406, $p = .046$, respectively.

Table 3. *Descriptive Statistics for Vocabulary Test Scores, Verbal SAT, and GPA For Each Condition*

Condition	Mean	N	SD	Min.	Med.	Max.
	Vocabulary Test Scores					
Match features	63.61	49	19.14	19	63	99
Explain steps	54.00	50	17.32	22	55	90
Solve only	61.46	50	21.22	10	59	99
	Verbal SAT					
Match features	585.94	32	93.39	420	585	760
Explain steps	505.67	30	96.37	330	505	680
Solve only	553.71	35	91.93	350	560	720
	GPA					
Match features	3.03	34	.648	1.53	2.93	3.97
Explain steps	2.67	34	.667	1.23	2.64	3.93
Solve only	2.72	35	.596	1.14	2.75	3.95

Materials

Sixteen problems pairs for training (one mixture, one distance) were constructed so that the members of the pairs could be either the same as one another, or different, in terms of their solution procedures (as explained in the introduction). As shown in Table 4, the test problems had either one of the old (trained) procedures or a new procedure. Test problems with both the same surface features as that used in training, and different surface features were also used. Crossing these factors resulted in four kinds of test problems. Two tests of equivalent difficulty, each test containing one each of the four kinds of test problems, were constructed. One was used as a pretest and one as a posttest, with the order counterbalanced across participants.

Table 4. *Examples of the test problems*

	Solution Procedure: Old	Solution Procedure: New
Surface features		
Old	A chemist combines 5 qts. of a 40% acid solution with 15 qts. of a 20% acid solution. What is the resulting % of acid of the whole 20 qts. of solution?	A wine company has 40 gallons of wine that is 25% alcohol. They need to combine it with some wine that is 4% alcohol, so that the resulting wine will be 10% alcohol. How much of the 4% alcohol wine should be added to the original 40 gallons of wine?
New	In a small town with a population of 25,000 people, there are 5,000 long time residents, and their average income is $40,000 per year. The other 20,000 people are newcomers, and their average income is $45,000 per year. What is the average yearly income of the entire town?	In an experiment on the effects of smoking on heart rate, the average heart rate for 35 male subjects in the experimental group was 78 beats per minute. The average heart rate for the female subjects in the experimental group was 76. The average heart rate for the entire experimental group was 76.5. How many female subjects were

Procedure

The procedure consisted of a pretest, a three part training session, and a posttest. In the pretest the participants were presented with one each of the four kinds of test problems in a random order. They were allowed three minutes to work on each problem. The posttest contained the same types of problems as the pretest and were of equivalent difficulty, although they were different versions of those problems. The posttest and pretest were counterbalanced across conditions.

The three-part training procedure consisted of worked examples, guided practice, and unguided practice. All participants were trained in the correct step-by-step solution procedures for the problems (but without any training in using a weighted averaging equation), and were also (if they were in one of the experimental conditions) trained to do their experimental tasks. They were trained with two pairs of worked examples, which were explained to them. Then they worked on two pairs of guided practice problems, in which they solved all of the problems, and did their training task with assistance from the experimenter who insured that they eventually did everything correctly. Then they worked unguided, again solving the problems and doing the training tasks, doing as many pairs as they could in fifteen minutes. The training tasks for the participants differed depending upon which condition they were in.

In the first experimental condition (N = 50), called the match features condition, the problems were presented explicitly as pairs. The participants matched those elements from the members of the pairs that corresponded in terms of being either initial ratios (or their associated amounts), or being the final, averaged ratio (or its associated amount). They did this by simply writing a list of pairs of elements matched according to initial and final elements. For some pairs (as explained in the introduction and as shown in Table 1), the matching task resulted in not only matching initials with initials and finals with finals (thus, resulting in matching them according to an analogy based on the principle of weighted averaging), but also matching the elements according to givens and unknowns (thus, resulting in matching them according to an analogy based on the solution procedure). Other pairs were like the pair shown in Table 2. One of the second initial elements from Problem A (the 10% of brown rice in the 100 lbs. of mixed rice), which was given, was matched to the unknown from Problem B (the speed of the plane for the last 3 hrs. of the trip), and one of the final elements from Problem A (the percentage of rice in the whole 250 lbs. of mixed rice), which was the unknown, was matched with one of the givens in Problem B (the 120 mph that second plane flew for the whole 5 hrs.). Importantly, however, even though doing the matching in this way did not result in matching according to an analogy based on the solution procedures (Problem B had a subtraction where Problem A had addition), it did result in matching according to an analogy based on the weighted averaging principle.

In the second experimental condition (N = 50), called the explain steps condition, the training problems were presented sequentially rather than explicitly as being in pairs. The participants were instructed to write an explanation for why each step of the solution procedure was necessary as they performed each step during solving. Explaining each step of what one is trying to understand during the process of trying to understand it is the way explaining was done in previous studies that showed positive effects of self-explanation. Therefore, the present participants explained in this way, but without associating one problem with another. Furthermore, the worked examples and the guided practice examples included a sample of an explanation for them to follow. This is consistent with the finding of Renkl, Stark, Gruber, and Mandl (1998) that modeling how to produce self-explanations of worked examples was superior to allowing participants to producer their own spontaneous self-explanations. Doing so improved posttest and far transfer performance (on problems with different underlying structures) especially for learners with low prior knowledge. An example of a sample self-explanation for one of our worked example problems can be seen in Appendix A. An example of a typical participant's self-explanation can be seen in Appendix B.

In the control condition (N = 50), called the solve only condition, the participants solved individually presented practice problems. The instructions that the participants were given were just the three steps of the solution procedure.

The participants were run in sessions with either a single participant, or a small group in each session. Each session was randomly assigned to a condition so that 50 participants ended up in each condition (not counting the 7 procedural errors). There were a total of 46 sessions consisting of 12 for the match features condition, 15 for the explain steps condition, and 19 for the solve only condition. Across the three conditions, the single participant sessions accounted for 28% of the sessions, sessions of from 2 to 6 participants accounted for another 63%, and the remaining 9% had either 7 or 8 participants. Those percentages were similar for each individual condition.

Design

The experiment had a 2 (order of test) by 2 (new procedure, old procedure) by 2 (new surface features, old surface features) by 3 (match features, explain steps, solve only) design.

Results

Performance During Unguided Practice

Given that during the unguided practice it was the time on task that was controlled, rather than the number of problems, it was possible for the number of problems completed to be different across conditions. Accordingly, a one-way analysis of variance showed that there were significant differences in the number of unguided practice problems completed, $F(2,146) = 105.12$, $p < .0001$ (the error degrees of freedom are one less than would be expected because, due to a clerical error, there was missing data for one participant). Scheffe all pairwise comparisons at the .05 significance level showed that all the means were significantly different from one another. As shown in Table 5, the solve only participants completed the most problems, followed by the match features and then the explain steps participants. The numbers of unguided practice problems solved correctly followed the same pattern, which resulted in the percentage of unguided practice problems solved correctly being not significantly different across conditions. The mean percent of unguided practice problems solved correctly by the match features, the explain steps, and the solve only participants were .83, .77, and .79, respectively.

Table 5. *Descriptive Statistics for Number of Unguided Practice Problems Completed for Each Condition*

Condition	Mean	N	SD	Min.	Med.	Max.
Match features	4.98	50	1.33	2	5	8
Explain steps	3.08	50	0.94	1	3	5
Solve only	7.71	49	2.25	3	8	12

Performance on the Posttest with Pretest Scores as a Covariate

As explained in the participants subsection of the methods section, there were significant differences between conditions on the individual difference variables of vocabulary test score, verbal SAT, and GPA. Therefore, all of the analyses reported in the paragraph below were repeated using all of those individual difference variables, and all possible combinations of them, as covariates in addition to the pretest scores. Those analyses showed that none of the results reported below changed from either significant to nonsigificant or from nonsignificant to significant as a result of using the additional covariates.

As shown in Table 6, on the posttest, for the new procedure problems only, the match features participants performed better than the explain steps participants and also (in spite of having solved fewer unguided practice problems) better than the solve only participants. The match features condition was superior to both the explain steps condition by itself and the solve only condition by itself, $F(1, 383) = 16.22$, $MSE = .14$, $p < .001$, $\eta 2 = .037$, and $F(1, 383) = 8.79$, $MSE = .14$, $p = .003$, $\eta 2 = .020$ respectively. However, the explain steps condition did not differ from the solve only condition, $F(1, 383) = 1.18$, $MSE = .12$, $p > .05$. Finally, the match features condition was also superior to the combined explain steps and solve only conditions, $F(1, 583) = 17.22$, $MSE = .13$, $p < .001$, $\eta 2 = .026$.

Table 6. *Mean Percent Correct on the Posttest (Adjusted for Pretest Performance) As a Function of The Procedure Required To Solve the Problems And the Type of Training*

	TrainingCondition		
Procedure	Match Features	Explain Steps	Solve Only
New	32	11	16
Old	70	66	64

For the old procedure problems only, there was no simple effect of training, across all three conditions, $F(2, 575) < 1$. There was no main effect of training condition collapsing across the old and new procedure problems, $F(2, 143) = 1.32$, $MSE = .17$, $p > .05$. There were no other main effects or interactions involving training condition, surface features, or procedure.

Participants could attempt to solve the problems by several strategies. They could solve them either by using the procedure in which they were trained, by using some other procedure, or (in spite of not being trained in itsuse) by using the weighted averaging equation. Therefore, in addition to being scored as correct or incorrect on each problem, the participants were also scored for use of the trained procedure and for use of the equation.

The new procedure problems were constructed so that it would be virtually impossible to solve them by any step-by-step strategy, and certainly it would be impossible to solve them by the step-by-step trained procedure. Rather, using the weighted averaging equation would be virtually required. As shown in Table 7, the match features participants used this beneficial strategy more than the other participants. The match features participants used the equation significantly more than the explain steps participants only, $F(1, 383) = 3.95$, $MSE = .10$, $p = .047$, $\eta^2 = .007$. Although their use of the equation was only marginally greater than that of the solve only participants alone, $F(1, 383) = 3.22$, $MSE = .10$, $p = .073$, $\eta^2 = .006$, it was significantly greater than that of the two other conditions combined, $F(1, 583) = 6.15$, $MSE = .09$, $p = .013$, $\eta^2 = .008$. On the other hand, there was no difference between the explain steps condition and the solve only condition, $F(1, 383) < 1$, n.s.

Table 7. *Mean Percent of Use of Trained Procedure and Use of Equation on the Posttest (Adjusted for Pretest Performance) As a Function of The Procedure Required To Solve the Problems And the Type of Training*

	Training Condition		
Procedure	Match Features	Explain Steps	Solve Only
	Use of Equation		
New	29	19	21
Old	32	17	16
	Use of Trained Procedure		
New	6	17	19
Old	87	77	88

Attempting to use the equation on the old procedure problems is not necessarily required, but it would nevertheless be beneficial for solving those problems. As shown in Table 7, the match features participants attempted this beneficial strategy significantly more than the other participants. The match features participants used the equation more than the explain steps participants only and more than the solving participants only, $F(1, 383) = 11.20$, $MSE = .10$, $p = .001$, $\eta^2 = .021$ and $F(1, 383) = 12.98$, $MSE = .10$, $p < .001$, $\eta^2 = .023$ respectively. They also used it significantly more than those two conditions combined, $F(1, 583) = 18.59$, $MSE = .09$, $p < .001$, $\eta^2 = .022$. On the other hand, there was no difference between the explain steps condition and the solve only condition, $F(1, 383) < 1$, n.s.

As explained above, attempting to use the trained procedure on the new procedure problems is a strategy that would be detrimental for solving those problems. As shown in Table 7, the match features participants attempted this detrimental strategy the least of all the participants. The match features participants used this detrimental strategy less than the explain steps participants only and less than the solving participants only, $F(1, 382) = 9.72$, $MSE = .06$, $p = .002$, $\eta^2 = .023$ and $F(1, 382) = 16.46$, $MSE = .05$, $p < .001$, $\eta^2 = .040$ respectively. They also used it significantly less than those two conditions combined, $F(1, 582) = 14.11$, $MSE = .07$, $p < .001$, $\eta^2 = .023$. On the other hand, there was no difference between the explain steps condition and the solve only condition, $F(1, 383) < 1$, n.s.

Attempting to use the trained procedure on the old procedure problems would be a beneficial strategy for solving those problems. As shown in Table 7, the explain steps participants used this beneficial strategy the least. The explain steps participants used it less than the match features participants only and less than the solving participants only, $F(1, 382) = 7.17$, $MSE = .06$, $p = .008$, $\eta^2 = .017$ and $F(1, 383) = 7.24$, $MSE = .09$, $p = .007$, $\eta^2 = .018$ respectively. They also used it significantly less than those two conditions combined, $F(1, 582) = 10.85$, $MSE = .07$, $p = .001$, $\eta^2 = .018$. On the other hand, there was no difference between the match features condition and the solve only condition, $F(1, 382) < 1$, n.s.

Discussion

The results of this experiment show that the process of matching features of algebra word problems is superior to either trying to explain their solution procedures, or merely practicing solving them, for facilitating people's ability to infer the general principle on which the problems are based. It is surprising that the present study showed that people benefited by feature matching not only compared to a control condition involving only solving, but also compared to a condition involving verbal explanation. Furthermore, they did so by engaging in a process that not only did not involve verbal explanation, but also was similar to the kind of pattern matching that is usually considered to be not beneficial for learning. It is especially impressive that the feature matchers applied the general equation for weighted averaging in situations where it was called for, given that they received no training in using the equation. It is equally impressive that they avoided an inappropriate use of the trained procedure on the new procedure problems. Being able to make a distinction as to when to apply and when to not try to apply a trained procedure is indicative of the kind of understanding that educators seek. Apparently, just the fact that the matched elements are related to the principle, enables people to induce important and useful information about the principle, without making that information explicit. Also, the result cannot be attributed to any difference in the number of problems completed during the unguided practice, because the match features participants actually completed fewer problems than one of the groups of participants (the solve only participants) that they outperformed on the posttest.

It might be noted that the feature matchers, in addition to being the only participants in the study who engaged in the feature matching task, were also the only participants for whom the members of the pairs of training examples were explicitly presented simultaneously (i.e., side by side). For the other participants, the same examples were presented sequentially. This raises the possibility that the positive effects of feature matching could be partly due to other consequences of seeing the examples simultaneously, such as having the opportunity to compare and contrast the examples in ways that the other participants did not, or to diagrammatically map the one-to-one correspondences between the problem pairs. In a previous study using both the same types of problem pairs as in the present study and some other types of problem pairs (described below), but always presented simultaneously, Craig and Ryan (2010) failed to find a positive effect of feature matching. On the one hand, they found that feature matchers did not perform significantly better on transfer problems than participants who were simply instructed to make a free comparison (i.e.., to write down whatever they saw was similar or different about the two problems, but with no training in matching the features). However, Craig and Ryan's transfer problems had only new surface features, not new solution procedures, and it was on the new procedure transfer problem in the present study that we did find a positive effect of feature matching. Also, unlike in the present study, in which members of the training pairs always had the same underlying principle of weighted averaging of rations, in Ryan and Craig's study, only half of the participants received such training pairs. The other half received training pairs in which one member was based on that underlying principle, and the other member was not. Finally, the sample sizes in the conditions in Craig and Ryan's study were from N=37 to N=40, as opposed to the N=50 used in the present study. Therefore, a direct test of the possibility that simultaneous presentation, rather than feature matching accounts for the results reported here is still needed. The best test would be to do a direct replication of the present study except to cross the training conditions used in the present study with simultaneous versus sequential presentation of the training pairs.

Regarding diagrammatic mapping, Gick and Holyoak (1983) found that such a process supports analogical transfer. Therefore, a direct test of whether the findings in the feature matching condition are attributable to feature matching or diagrammatic mapping is still needed. Such a test might be implemented by, for example, comparing conditions in which participants either stated verbally which features matched or drew a diagram connecting the matching features.

In addition to replicating Ryan (2005), the results of the present study have theoretical implications for understanding self-explanation, and practical implications for instruction in mathematics. Algebra teachers might consider verbally explaining how an algebra word problem was solved as a useful learning activity, especially in light of the research on self-explanation. Indeed, there are probably many situations in which it is useful. The present research, however, suggests caution regarding the use of self-explanation.

Although many previous studies have found self-explanation to be useful for learning and transfer (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, DeLeeuw, Chiu, & LaVancher, 1994; Pirolli & Bielaczyc, 1989), it is important to consider that self-explanations can be of different types, and that their usefulness varies. Renkl (1997) examined the various types of self-explanations produced by students, and the relation between type of explanations produced and learning outcomes. Renkl found that various types of self-explanations lead to successful learning outcomes, while others lead to unsuccessful learning outcomes. Participants who explained underlying principles performed better on a posttest, and on both near and medium transfer. For our participants, the principle that division is the inverse operation of multiplication was relevant for the last step of their problems. They were given a sample self-explanation that embodied that principle in the explanation of the last step (see Appendix A). Although we did not analyze their self-explanations, a quick look at them shows that they did not explain that principle. Renkl also found that making use of anticipative reasoning was positively related to better performance on the posttest, near transfer, and medium transfer. For our participants, using anticipative reasoning would have involved explaining something like, “In this problem you are looking for a final percentage, not an initial percentage. Therefore, this problem will involve adding two initial amounts of constituent material, not subtracting one of the initial amounts from the total.” Our participants did not engage in such anticipative reasoning, although, in their defense, our sample self-explanation did not include it either. Another type of self-explanation identified by Renkl was combining goals and operators. The sample self-explanation illustrated that type for our participants in all three steps of the problem (see Appendix A). Our participants engaged in this type of self-explanation to a great extent (see Appendix B). However, that is only one of three types of self-explanation that Renkl found to be beneficial, and it is one that was beneficial for performance on the posttest and for medium transfer, but not for near transfer. Thus, our participants engaged in only one of several possible useful types of self-explanation that they could have used, and it was not one of the most useful. In summary, if self-explanation is to be considered as a possible instructional method for algebra learning, it may be important to be sure to elicit the most useful types of self-explanations. Examining that question further remains for future research.

However, if the educational goal is to help students learn an underlying principle such as weighted averaging of ratios and then to transfer it to problems in which they had to recognize that the procedure in which they were trained did not apply, and to generate a new procedure based on the principle, then a task that involves forming a principle based analogy between example problems might be useful. The match features task used in the present experiment produced such a benefit even though it could be seen as mindless pattern matching. In fact, the superiority of the feature matchers over the other participants in the present study was even slightly greater than the superiority of the similarity judgers in Ryan (2005).

Although there are benefits from problem comparison tasks there are costs as well. To some extent with either feature matching or with the similarity judging task used in Ryan (2005), but especially with the feature matching task, a hurdle to be overcome would be to devise a task that would accomplish the feature matching benefit, but which would appear meaningful to students. Engaging students' interest and grounding their learning in ways that they find personally relevant is a challenge for teaching in any domain. However, in teaching mathematics, the relevance of which many students question, such a concern is only that much greater.

Students who have little to no personal stake in the task often have little motivation to perform well, and it has been found that this low motivation is correlated with below average performance on tasks (Cole & Osterlind, 2008). However, when otherwise low-stakes assessments were presented as having a consequence designed to increase motivation, such as telling students that their test scores may be seen by faculty in their department or by potential employers to evaluate their academic performance, their test scores were significantly better (Liu, Bridgman, & Adler, 2012). Therefore, when using feature matching as an actual instructional method it may be necessary to present it as involving a consequence that is meaningful to students, which in secondary education might be something such as making it an assignment to be graded.

Also, in keeping with the findings about the types of self-explanations that are most useful for learning from comparing examples (Renkl, 1997), a good instructional method might involve developing techniques to elicit such self-explanations. Those techniques could involve either just instructions and examples, or modeling the generation of useful self-explanations. In any event, it may be worth the effort to develop instructional methods that exploit the benefits shown in this study. This is especially true because the method would involve teaching with examples.

Teaching principles with examples capitalizes on people's preference for examples over abstract instruction (LeFevre & Dixon, 1986) but, in some cases, runs the risk of producing some negative transfer to new problems that require a new procedure (Luchins, 1942; Ryan & Schooler, 2001). However, the present study shows that the feature matching task actually reduced the risk of negative transfer. It produced significantly less tendency to mistakenly attempt to use the procedure from the training examples on the problems that required applying knowledge of the principle to develop a new procedure. Therefore, an instructional method that was based on forming an analogy between the principle based structure of example problems, if designed to appear meaningful and relevant to students, could be very useful for teaching principles such as the weighted averaging of ratios.

AUTHOR NOTES

Correspondence should be addressed to Robert Ryan; [email:rryan@kutztown.edu]. The authors wish to thank Jerry Casale, Barb Pentz, and Cheryl Smith for help in collecting the data. We thank Jonathan Schooler for comments on earlier versions of the manuscript. We thank Gretchen Kunkel for coding, and scoring the strategy use data. And, finally, we thank two anonymous reviewers for their helpful comments on a previous version of this manuscript.

REFERENCES

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M. T. H., DeLeeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Craig, S. J., & Ryan, R. S. (2010, May). Aptitude is better than number of analogical training examples for predicting problem solving performance. Paper presented at the annual meeting of the Association for Psychological Science, Boston, MA.

Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science, 15,* 89-132.

Cole, J. S., & Osterlind, S. J. (2008). Investigating Differences between Low- and High-Stakes Test Performance on a General Education Exam. *Journal of General Education, 57*(2), 119–130.

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

Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology, 25*, 524-575.

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

LeFevre, J. & Dixon, P. (1986). Do written instructions need examples? *Cognition and Instruction, 3*, 1-30.

Luchins, A. (1942). Mechanization in problem solving: The effect of Einstellung. (Psychological Monographs, Vol. 54, No. 6 [whole no. 248]). Evanston, IL: American Psychological Association.

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Educational Researcher, 41*(9), 352-362.

Pirolli, P. L., & Bielaczyc, K. (1989). Empirical analyses of self-explanation and transfer in learning to program. Proceedings of the 11th annual conference of the Cognitive Science Society (pp. 450-457). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Reed, S. K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 106-125.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*(1), 1-29.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23*, 90-108.

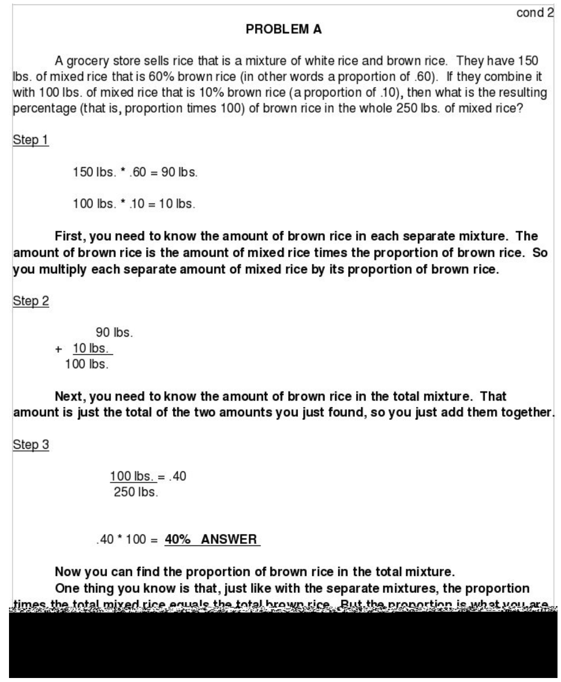
Ryan, R. S. (2005). Judging similarity facilitates deriving a new solution procedure. *Cognitive Technology, 10*(2), 5-12.

Ryan, R. S., & Schooler, J. W. (2001, June). Analogical reasoning with examples facilitates inferring general principles in spite of the risk of negative transfer. Paper presented at the annual meeting of the American Psychological Society, Toronto, ON.

.

Appendix A

An Example of the Sample Explanations (in Bold Face) Provided to the Participants



cond 2

PROBLEM A

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 (in other words a proportion of .60). If they combine it with 100 lbs. of mixed rice that is 10% brown rice (a proportion of .10), then what is the resulting percentage (that is, proportion times 100) of brown rice in the whole 250 lbs. of mixed rice?

Step 1

150 lbs. * .60 = 90 lbs.

100 lbs. * .10 = 10 lbs.

First, you need to know the amount of brown rice in each separate mixture. The amount of brown rice is the amount of mixed rice times the proportion of brown rice. So you multiply each separate amount of mixed rice by its proportion of brown rice.

Step 2

90 lbs.
+ 10 lbs.
100 lbs.

Next, you need to know the amount of brown rice in the total mixture. That amount is just the total of the two amounts you just found, so you just add them together.

Step 3

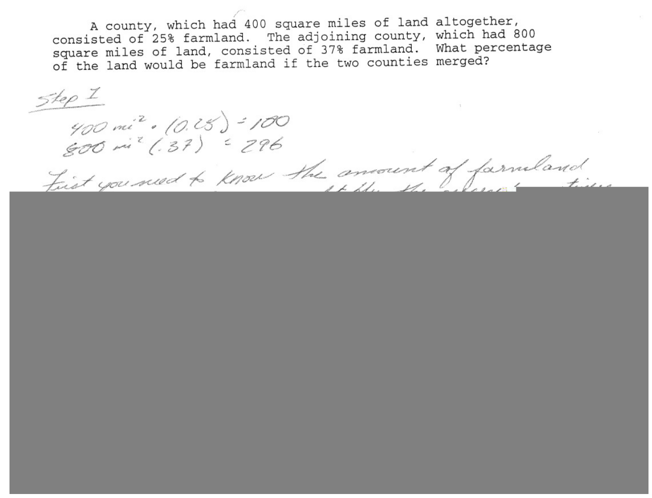
100 lbs. / 250 lbs. = .40

.40 * 100 = **40% ANSWER**

Now you can find the proportion of brown rice in the total mixture.
One thing you know is that, just like with the separate mixtures, the proportion
[redacted]

Appendix B

An example of a Typical Participant's Self-Explanation



A county, which had 400 square miles of land altogether, consisted of 25% farmland. The adjoining county, which had 800 square miles of land, consisted of 37% farmland. What percentage of the land would be farmland if the two counties merged?

Step 1

400 mi² • (0.25) = 100
800 mi² (.37) = 296

First you need to know the amount of farmland

.