Using Similarity Rating Tasks to Assess Case Reuse in Problem Solving

Fran Mateycik[1], David H. Jonassen[2] and N. Sanjay Rebello[1]

[1]*Department of Physics, 116 Cardwell Hall, Kansas State University, Manhattan, KS 66506-2601*
[2]*Department of Educational Psychology, 221C Townsend Hall, University of Missouri, Columbia, MO 65211*

Abstract. Case-reuse strategies involve extracting the conceptual schema from previous cases and adapting them to new problems. Recognizing the deep structure differences and similarities between problems is essential for productive case reuse. We report on a semester-long study with students participating in weekly focus group learning interviews to facilitate case reuse strategies. At the mid and end points of the study, students were interviewed individually to ascertain the effect of these strategies. During these interviews students were asked to rate the similarities between problem pairs. We report on the results from the similarity ratings as well as present a comparison with expert responses to these questions.

Keywords: problem solving, algebra-based physics, physics education research, focus groups, similarity ratings
PACS: 01.40.Fk

INTRODUCTION

Problem solving is an important cognitive skill that all people, especially those in STEM disciplines must develop [1, 2]. Our overarching study focuses on case reuse, a process of solving problems by using what was learned through similar previously solved problems [3]. Case reuse is based on the premise that students construct a conceptual schema by analyzing a worked example and retrieve this schema while solving similar problems. For a schema to be useful in problem solving it must be tied to the deep structure of the problem rather than its surface features.

Our overall goal was to facilitate the development of conceptual schema by enabling students to focus on deep structure of problems. During this project, we looked at assessing whether our treatment affected students' identification of deep-structure similarities between problems using a similarity rating task. Students participating in our study were given the task of rating the similarity between pairs of problems of varying similarities in surface features and deep structure features.

We address the following research questions:

Q1) To what extent do students attend to the surface or deep-structure features in comparing problems?

Q2) How do students' ratings of similarity between problems compare with faculty members' ratings of the same problems?

Research suggests that learners fail to recall examples or schema appropriately because their retrieval is based upon surface similarity between cases, not their deep structural features [4, 5, 6]. Catrambone and Holyoak also suggest that generalization improves when problems emphasize structural features shared with a similar example. Research by Chi [6] has shown that students tend to group problems based on surface features, while experts group problems based on their deep structure. Our tasks were different from those presented by Chi in her research. Rather than ask students to categorize the problems we presented students with pairs of problems and asked them to rate the similarity of each pair on a five-point Likert scale with '0' labeled as 'completely different' and '5' labeled as 'identical.'

METHODOLOGY

Ten students participated in the eight, 75-minute long, focus group learning interview sessions. These students were representative of the class demographic profile. The topic in each session followed those currently being covered in the algebra-based physics class all participants were enrolled in.

During each session a moderator handed out a fully solved example problem and a pair of problems for students to work on. The example problem provided was comparable in physical principle to the unsolved problems. Participants worked in pairs with each student working on a different problem. All problems shared deep structure similarities but had surface differences. After students had solved these problems, they were asked to discuss their solutions with their partner briefly and discuss the similarities and differences between each of the problems.

To assess the impact of using direct deep-structure similar problem comparison during the group learning interviews, the students were also required to participate in two individual interview sessions, one toward the middle and the other toward the end of the semester. One of the individual interview tasks asked students to rate the similarities between pairs of problems. The problem pairs were constructed from problems that had facial (i.e. surface) similarities and differences as well as principle (i.e. deep structure) similarities and differences. The overarching concept remained the same across all problem sets, thereby leaving some basic similarity among all problems.

All four combinations of facial/principle similarities/differences were created. These are labeled problem pair types A, B, C and D in Table 1.

TABLE 1. Problem pairs for the similarity rating task

	Facial Similarity (**FS**)	Facial Difference (**FD**)
Principle Similarity (**PS**)	**A**	**B**
Principle Difference (**PD**)	**C**	**D**

Each student was presented with eight pairs of problems. Students were presented the problem pairs in order A, A, B, B, C, C and D, D. Students were not allowed to backtrack and change their similarity rating for any pair until the end of the sequence when they were given the opportunity to review their ratings for all pairs and decide whether they wanted to revise any of the similarity ratings.

Figure 1 below shows examples of the similarity rating tasks used in the study in Interview 1.

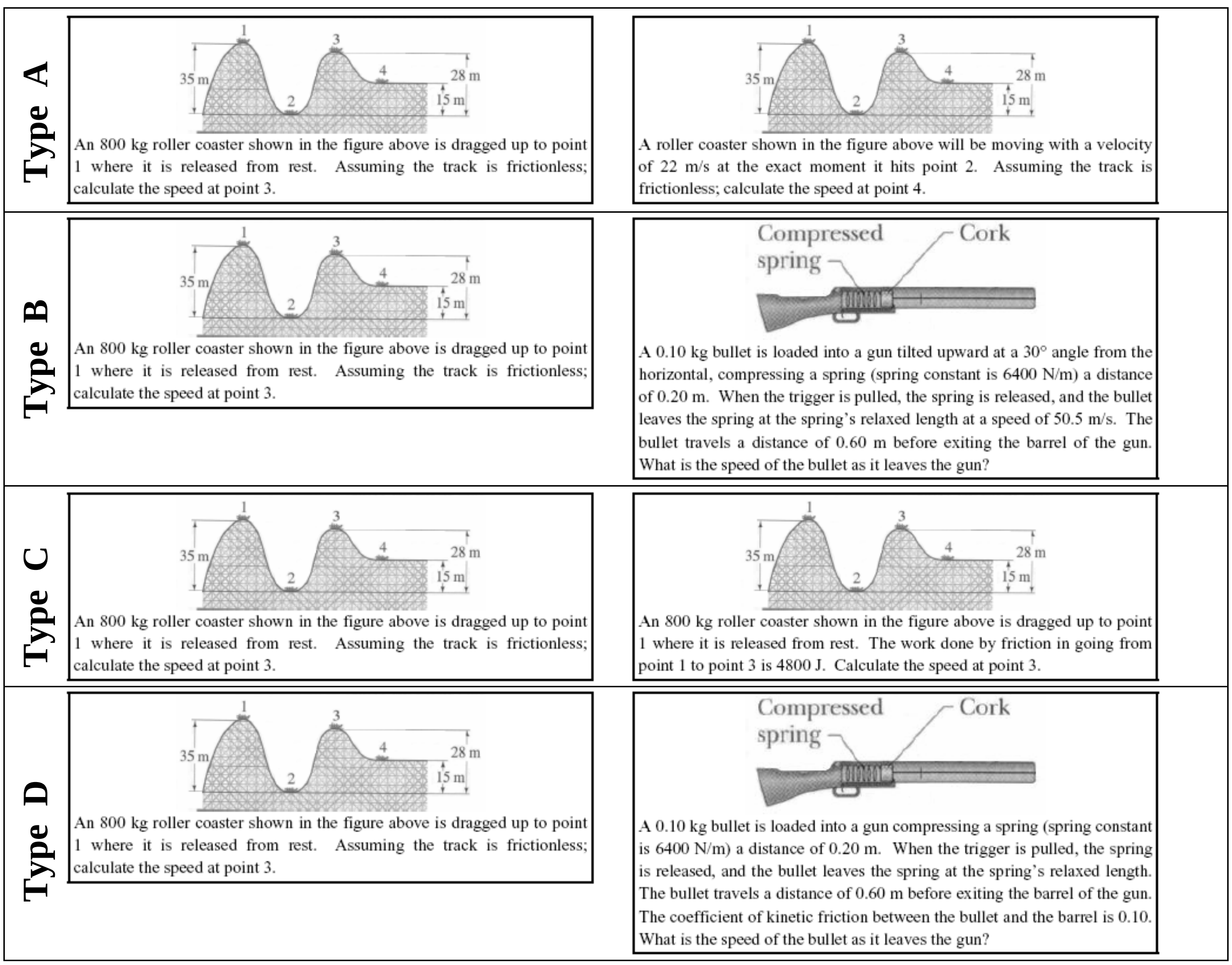

FIGURE 1. Examples of each type of problem pair used in interview 1.

Four non-PER (Physics Education Research) faculty members were also asked to complete the similarity ratings task at the end of the semester. All of these faculty members were either currently teaching or had recently taught an introductory physics course. PER faculty were asked not to complete this rating due to their familiarity with this project. Data were collected from these faculty members to compare with student data on similarity ratings. We expected the faculty members to be most sensitive to the principle similarities and differences, rather than facial similarities and differences. Thus, we expected that the faculty members would rate problem pairs A and B as 'high' on the Likert scale since they both shared principle similarities while rating pairs C and D 'low' on the Likert scale, because they both had principle differences.

RESULTS & DISCUSSION

We averaged the similarity ratings of each student for each problem pair type for each interview. Figure 2 below shows the rating for all four pair types for the first as well as second interviews.

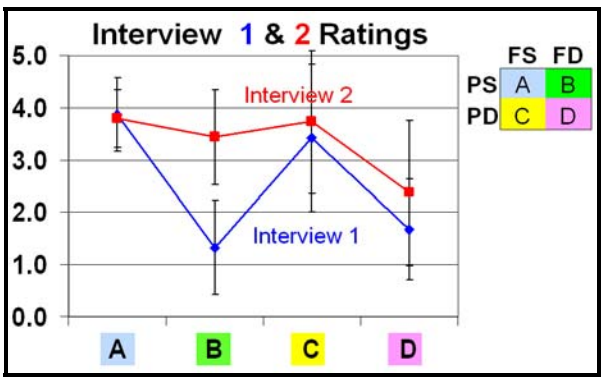

FIGURE 2. Students' similarity ratings of problem pairs of types A, B, C and D for interview 1 and interview 2.

Figure 2 shows principle/facial similarities/differences in each type (P=Principle, F=Facial, S=Similarity, D=Difference). The error bars are the standard deviation over all students and all problem pairs of a given type.

Interview 1 was conducted after students completed the first four focus group learning interview sessions. However, the protocols for these interviews were not finalized until the fourth interview, so students were not participating in activities that required them to explicitly focus and reflect on problem similarities and differences.

In our results for interview 1 we find statistically significant differences between the similarity ratings of pairs A and B (p-value 0.000), B and C (p-value 0.003) and C and D (p-value 0.008). The fact that students have rated pairs B and D as significantly lower than pairs A and C is consistent with the notion that students appear to be focusing on facial similarities and differences rather than similarities and differences in principle. For instance, they rate pair B significantly lower than pair A even though the problems in pair B are only facially different. Similarly, they rate pair C significantly higher than pair D even though the problems in pair C have differences in underlying principle.

Through discussion of the similarity ratings with students during this task, it becomes apparent that students recognize problems are related by conservation of energy, but they believe the differences in facial features have a direct effect on the types of energies involved, and these are enough to make the solution that much more different.

> *"I guess that both the stone and the piano have potential energy like when they're starting, but that doesn't matter really. It's a totally different technique used to solve each problem. There's a spring energy now."*

It is also apparent through the conversation that Pair C problems are different in terms of the method necessary to solve the problems, but are not 'significantly' different.

> *"Except this one you're gonna be using a tiny different equation in the path [solving procedure] than this one and that [part of the solution] was the same."*

Interview 2 was conducted after students completed all eight of the focus group interview sessions. At this point, students participated in five finalized focus group learning interviews.

Here we find that the differences between A and B, B and C are no longer statistically significant. The only statistically significant difference is between C and D (p-value 0.014). The fact that students are rating pairs A and B at about the same level of similarity is consistent with the notion that students have now begun to recognize that the problems in pair B have principle similarities that overpower their facial differences to the extent that they rate pair B almost the same way as they rate pair A. In other words, it appears from these data that students are emphasizing the similarities in principle although there may be facial differences between the problems in pair B. The ratings for pairs C and D in interview 2 are close to identical to their ratings for these pairs in interview 1. We would be interested in seeing the rating for pair C to be significantly less than before, and as low as the rating for pair D. Such data would have been consistent with the notion that students are able to overlook the facial similarities in pair C and recognize the difference in principle, but our data do not appear to show this pattern. Rather, it appears from our data that when shown a problem pair that is facially similar, students do not probe further to reflect on whether or not these problems' similarities/differences in principle are significantly impacting the solution.

Data collected from four faculty members at the same institution were compared to the data collected from students in interview 2. There were not enough faculty to warrant any statistical calculations, but we can see from the small sample that those faculty participants agree with the ideal hypothetical expert. Problem pair types A and B are both rated high and close to one another, while problem pair types C and D rate lower and close to one another. Figure 3 below shows the average rating for each problem pair type given by faculty.

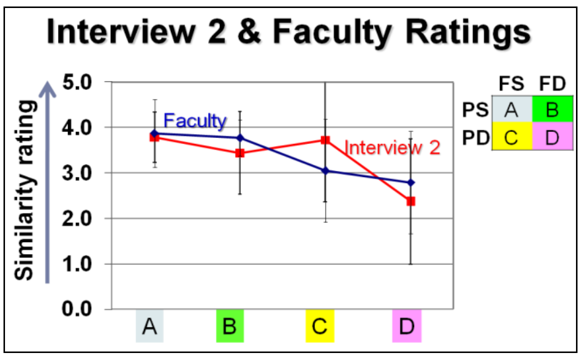

FIGURE 3. Student ratings from interview 2 are mapped on top of faculty ratings for comparison.

It can also be seen in Figure 3 that students' ratings for three of the four problems are similar to the faculty' ratings by the end of the semester. Problem type C is most different. Students rate type C problem pairs higher than type A and B problem pairs, while faculty rate type C problem pairs lower than type A and B problem pairs.

CONCLUSIONS

We address each of our research questions below:

Q1) To what extent do students attend to the surface and deep-structure features in comparing problems?

Before our focus group learning interviews, students rated problems sharing prominent surface features higher than problems with different surface features. After our focus group learning interviews, students' ratings of problems sharing surface features remained high, but problems with different surface features and similar deep-structure features were also rated high.

Q2) How do students' ratings of similarity between problems compare with faculty members' ratings of the same problems?

A direct comparison is difficult with such small numbers, but if we look at general trends, the faculty' ratings and students' ratings after treatment are very close for problem pair types A, B and D. Problem pair C, which includes problems that are facially similar and principle different, are rated lower than problem pair types A and B for faculty, but higher than problem pair types A and B for students. Students learn to deemphasize facial features when given problems that are not facially similar. When problems share facial similarity, the students no longer attend to the differences in principle between problems.

LIMITATIONS & FUTURE WORK

The change in students' ability to discern the similarities and differences in interviews 1 and 2 could be due to not only the participation in the focus group learning interviews. They could also be due to the differences in the specific problems used in each interview and/or the topic on which they were based. Students were simultaneously enrolled in an algebra-based physics course which also could have altered the deep-structure feature emphasis on these problem similarity rating tasks. Future work would need a larger experimental sample, a control or baseline group, and/ or a wider variance of student population. The similarity variance between problem pairs could also be augmented to include greater differences in structure and problem representation.

ACKNOWLEDGMENTS

This research is supported in part by NSF grant DUE-06185459.

REFERENCES

1. L. Hsu, E. Brewe, T.M. Foster, and K.A. Harper, *American Journal of Physics* **72**(9), 1147-1156 (2004).
2. D.H. Jonassen, *Educational Technology and Research and Development* **48**(4), 63-85 (2000).
3. B. Faltings, *ICCBR Proceedings*: 611-622 (1997).
4. R. Catrambone and K.J. Holyoak, *Journal of Experimental Psychology* 15(6) 1147-1156 (1989)
5. S.K. Reed and C.A. Bolstad, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 753-766 (1991)
6. M.T.H. Chi, P.J. Feltovich, and R. Glaser, *Cognitive Science* **5**(2), 121-152 (1981).