FACILITATING CASE BASED REASONING IN PHYSICS PROBLEM SOLVING

Case-based reasoning (CBR) is the process of solving a real-world problem based on precedent examples and problems. A cohort of 10 students participated in focus group learning interviews in an algebra-based physics class eight times during the semester. Participants worked in pairs to solve and discuss problems that shared deep structure similarities and surface differences. We collected data on non-traditional problems inserted into each of five multiple choice examinations during the semester. Our results show a statistically significant difference on some of the non-traditional questions between our cohort group and the rest of the class. Additionally, these cohort students were interviewed individually at the mid-point and the end of the semester during which they were asked to rate the similarity between problem pairs. Students' problem similarity ratings in the second interview compare favorably in some ways to the first interview, but important issues remain to be addressed.

Fran Mateycik, Kansas State University
N. Sanjay Rebello, Kansas State University
David H. Jonassen, University of Missouri, Columbia

Introduction

One of the important issues in problem solving literature is the differences between experts and novices in their problem solving strategies. Maloney (Maloney, 1993) points out several issues that distinguish novices from experts in the ways in which they approach problem solving. Some of them include use of the 'means-ends' analysis to solving problems, focus on quantitative procedure rather than qualitative reasoning, and inability to recognize when they may have solved a problem incorrectly. Several strategies to remedy these issues have been suggested in the literature. Hsu (Hsu, Brewe, Foster, & Harper, 2004) provides an excellent review of problem solving literature in physics education research.

In this paper we describe a research study that focuses on one particular strategy – case reuse – to enable students to improve their problem solving skills and transition from novice to expert-like problem solving strategies. Research (Jonassen, 2006) has shown that case reuse strategies hold some promise in enabling students to reflect on and extract the conceptual elements of a case, such as a solved problem, to improve their problem solving strategies, particularly their attention to the conceptual schema underlying the problem rather than the procedural schema, which most novices tend to focus on.

Our goal is to facilitate the development of conceptual schema during problem solving using case-reuse strategies that help students focus on deep structural properties of a problem rather than surface differences. To achieve this goal we have conducted group learning interviews with a cohort group of students enrolled in an algebra-based physics course. We then assessed the impact of our intervention on students' conceptual schema using non-traditional problems on exams and similarity rating problems in individual interviews at the mid-point and toward the end of the semester.

We attempted to answer the following research questions in our study:

- How do students determine whether a given example is useful for solving a different problem?
- How might we refocus a student's emphasis on the similarities and differences between problems to include emphasis on deep structure differences?
- To what extent does the treatment facilitate students' development of conceptual schema as assessed by non-traditional problems and similarity rating tasks?

Relevant Literature

Case-based reasoning (CBR) may be generically defined as the process of solving a real-world problem based on analogies (Kolodner, 1997). CBR is not a set of procedures that carry out analogical reasoning. Rather, CBR suggests a cognitive architecture, or synthetic model of analogical reasoning, that integrates our natural reasoning skills with computational processing (Kolodner, 1997). In other words, once a case is retrieved, an old solution might be adapted to solve a new problem or several pieces from some old situations might be merged and applied to the retrieved case. The important issue is to focus on what needs to be adapted in the new case so that the learner may extract and merge elements from previous cases, and thus come away with something in memory that can be used to process the problem.

Case reuse is a strategy that promotes CBR (Kolodner, 1997) by employing problem pairs that share similarities in deep structure. In cognitive psychology, case-reuse refers to the process of solving problems based on analogy (Faltings, 1997). More recently, Jonassen (Jonassen, 2006) presents case reuse as a strategy in which learners are presented with problem solving cases as examples or analogues of how similar problems are solved. Students construct schemata, or mental representations, by analyzing a worked example. Schema may be retrieved as learners work solutions to new similar problems. The schema consists of knowledge about problem type, structural elements (acceleration, velocity, distance, etc.), situations in which the problem occurs (car, on an inclined plane, baseball, etc.), and the processing operations required to solve the problem. Research suggests that learners fail to recall examples or schema appropriately because their retrieval is based upon similarity of objects between examples, not their structural features (Catrambone & Holyoak, 1989; Reed & Bolstad, 1991). Catrambone and Holyoak also suggest that generalization improves when problems emphasize structural features shared with a similar example, and the number of examples is increased (i.e., three examples are better than two).

Methodology

We conducted studies on problem solving in an algebra-based physics course. Participants included 10 students that were randomly selected from the original 46 volunteers. The participants met in a single group a total of eight times during the semester for what we called focus group learning interviews. Each of the eight focus group learning interview sessions was 75 minutes long. Additionally, the participants met with the moderator individually twice during the semester. The individual interviews were conducted at the mid and end points of the semester. Each individual interview was 50 minutes long.

Figure 1 shows the research methodology on a timeline beginning from left at the start of the semester to the right at the end of the semester.

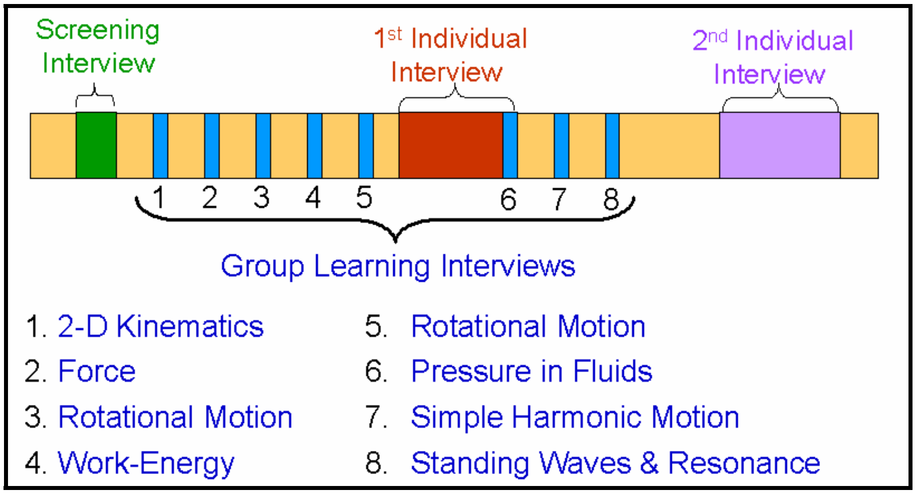


Figure 1. Research design timeline

Screening Interviews

Screening interviews, each lasting about 20-30 minutes were conducted with 21 participants that were selected from a pool of 46 volunteers. Each student was paid $8 for participating in the screening interview. The main purpose of the screening interviews was to gain insights about how students solved problems and whether or not they worked with others. Students were asked about the prior physics classes they had taken, including in college and high school. They were asked about their interest in the current physics class that they were taking as well as why they were taking it. Thus, we wanted to screen for students who were not apathetic toward the class or were very likely to drop out in the middle of the study.

We were most interested in selecting students who would be amenable to learning how to solve problems by looking at solved examples and also those who were comfortable working with others, since the group learning interviews were an interactive environment and we wanted to ensure that students who were selected in our study would be comfortable participating in it. To screen for these attributes, we asked students about their study habits, especially how they went about solving problems. A significant aspect of the study was case-reuse, whether they used solved examples, and if so how. Based on a previous study (Mateycik, Jonassen, & Rebello, 2008) we had seen that students tend to overly rely on using equations. We asked students if and how they would use equations in solving problems. Finally, we also asked students whether or not they found the textbook useful and if so, in what ways was the textbook useful to them.

Focus Group Interviews

A total of 10 volunteers were selected from our screening interviews. They were invited to participate in the focus group learning interview sessions. A total of eight focus group learning interview sessions were held during the semester – about one per week, except on weeks when students had exams and other commitments. The topics addressed in each group learning interview are also listed in Figure 1. These topics cover the typical topics that are covered in a first semester algebra-based physics course.

During each focus group learning interview session a moderator would hand out a pair of problems for students to work on. These problems were labeled problem A and problem B. Participants were paired together such that one would be asked to work on problem A while the other worked on problem B. The problems shared deep structure similarities but had surface differences. An example of the problem pairs used in the first interview is shown in Figure 2. Both of the problems present contrasting cases and are focused on the same physical principle (Newton's II Law), but have many surface differences such as vertical versus horizontal orientation and the different kinds of objects (blocks versus train cars) in the two problems.

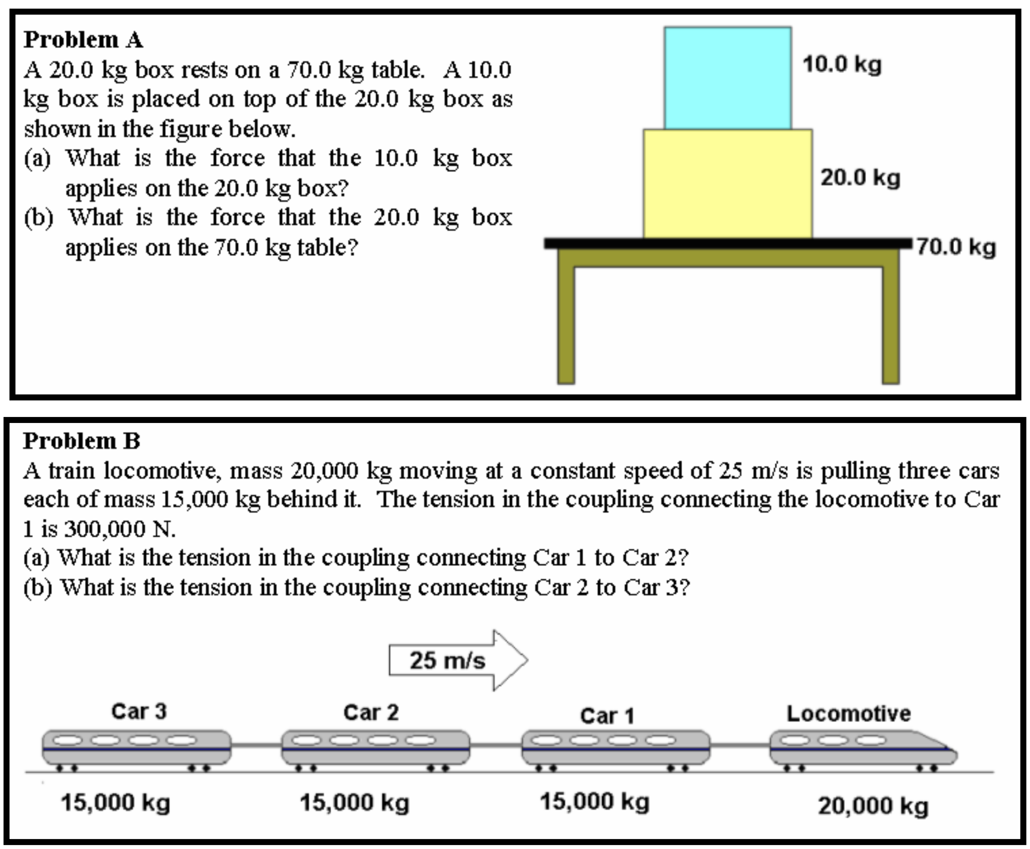
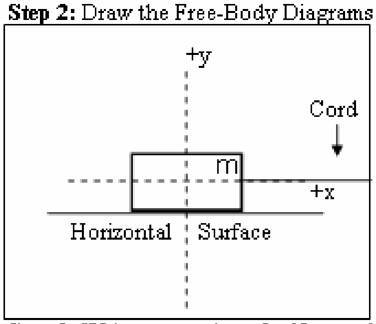


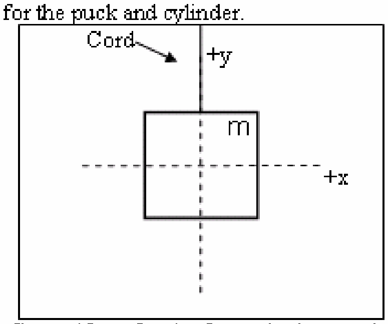
Figure 2. Example of contrasting cases presented to students in the group learning interviews. This is an example from week 1 of the sequence.

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. Finally, students were asked to work with their partner to create their own problem which uses some elements from both problem A and problem B.

In the first week of the semester, students often struggled with completing their individual problems. They had difficulties solving the problem and therefore did not have the time to engage in problem comparisons and discussions. To alleviate this difficulty, in the second week we introduced specific stopping points in the process at which students were asked to stop, signal to the facilitators and check with them about their progress in solving the problem and in the third week we provided more procedural scaffolding in the form of a more stepwise process of how to solve the problem as shown in Figure 3. The purpose of the procedural scaffolding was to decrease the cognitive load of the students in following the particular steps to solve the problem, so that they would be able to attend to the conceptual aspects of the problem such as reflecting on the underlying principles, similarities, and differences between problems.

Step 1: Identify and interpret the principles involved in the problem.
Step 2: Draw the Free-Body Diagrams for the puck and cylinder.





Step 3: Write expressions for Newton's Second Law for the forces in the x and y direction for each object.
Puck m: x direction

Puck m: y direction

Cylinder M: x direction

Cylinder M: y direction

Step 4: Specify whether the problem statement provides sufficient, missing, or irrelevant information for applying toward a solution.

Step 5: Use the principles and their respective equations identified above to solve the problem.

Step 6: Explain your solution to your partner.

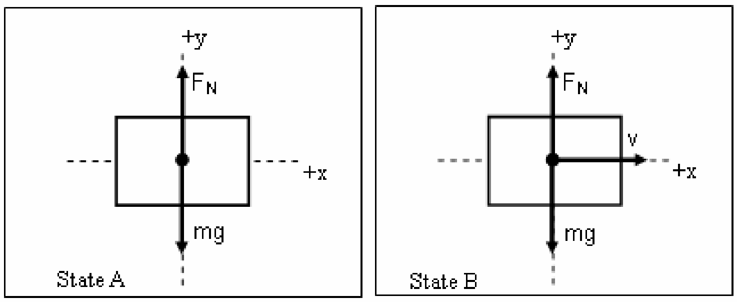
Figure 3. Example of the procedural scaffolding provided to students in the form of a step-by-step procedure. This particular worksheet was used in week 3 when students were asked to draw the free-body diagrams for the object.

Our experiences in the first three weeks of the semester taught us that students had difficulties in solving the problems. This meant that they had no time to reflect on the problem principles, similarities, and differences.

In week 4 students were given a worked example problem at the beginning of the group meeting. Research has shown that providing students with appropriate worked examples can facilitate problem solving (Ward & Sweller, 1990). The worked example problem, or problem C, included a full solution and was available as a resource for students to use while solving their own respective problems. Problem C would be deep structure similar to both problem A and problem B. Figure 4 shows an example of the solved example

(Problem C) followed the contrasting pairs (Problem A and Problem B). The facilitators went over Problem C in the first 15 minutes and then asked students to solve Problems A and B.

Problem C
Joshua pushes a 1.85 kg box along a flat horizontal table applying an average force of 39.0 N. The box starts at rest and reaches a velocity of 12.0 m/s. Neglecting friction, how far did Joshua push the box?



SOLUTION

We may express the work done by Joshua on the box in terms of the force applied and the distance covered while Joshua applied the force.
We know that since the box will be moving in the same direction as the force applied, the angle between the direction of force and direction of

$$W = Fd\cos(0) = Fd$$
$$\Rightarrow W = Fd = (39.0N)d$$

We also know that the work done on the box must be equal to the change in kinetic energy. Since the box was at rest initially our initial kinetic energy will be zero.

$$W = KE_2 - KE_1 = \frac{mv_2^2}{2} - \frac{mv_1^2}{2}$$
$$\Rightarrow W = \frac{mv_2^2}{2} - \frac{m(0)^2}{2}$$
$$\Rightarrow W = \frac{mv_2^2}{2}$$
$$\Rightarrow W = \frac{(1.85kg)(12.0m/s)^2}{2} = 133.2J$$

Finally, we have expressed the work done on the arrow using two different equations above. We may set both expressions for Work equal to one another

$$W = (39.0N)d = 133.2J$$
$$\Rightarrow d = \frac{133.2J}{39.0N} = 3.42m$$

Problem A
A 0.10 kg arrow is fired from a bow. The bow is pulled back a distance of 0.8 m so that the arrow is released with a speed of 50 m/s as it leaves the bow. The arrow travels 25.0 m before hitting its target. What is the average force exerted on the arrow by the bowstring?

Problem B
A Yankees batter hits a 0.14 kg baseball sending it off into left field, 40 m away from the batter's box. The baseball lands in a Royals fielder's glove, exerting an average force of 300 N, moving the glove backward 0.25 m before coming to rest. What is the speed of the ball just before it is caught?

Figure 4. The solved example (Problem C) that was presented to the students in week 4.followed by unsolved contrasting cases (Problem A and Problem B).

Based on our own observations of student performance in week 4 we realized that the protocol we had developed was successful in enabling students to work through the problems without significant barriers. The solved example (Problem C) gave students adequate scaffolding to complete the unsolved problems (Problem A and Problem B). We asked students to specifically score the usefulness of Problem C in solving Problem A and B. We also asked students to rate the similarities and differences between Problems A, B, and C. This protocol allowed adequate time for reflection and discussion after students had solved the problems. We continued with this protocol for the rest of the semester (i.e. week 5 through week 8), the only difference was that rather than go through Problem C at the beginning of the group learning interview, as we did in week 4, we asked students to go over Problem C first and ask us any questions that they had about it. After they had gone through Problem C, we provided then Problems A and B.

Individual Interviews

We conducted two individual interviews with all of the students in our focus group learning interviews. As shown in Figure 1, the first individual interview was conducted after week 4 of the focus group learning interview and the second individual interview was conducted at the end of the semester after all of the focus group learning interviews had been completed.

The purpose of these individual interviews was to assess the extent to which students' conceptual schema with regard to problem solving had evolved due to their participation in the focus group learning interviews. During the individual interviews students were asked to rate the similarities between contrasting problems of varying deep structure and non-deep structure similarities. Research by Chi (Chi, Feltovich, & Glaser, 1981) has shown that students tend to group problems based on surface features, while experts group problems based on their deep structure. Similarly, Hardiman (Hardiman, Dufresne, & Mestre, 1989) showed that surface similarities between problems could interfere with experts' classification of the problems. 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. Each student was presented with eight pairs of problems. The problem pairs of were constructed from problems that had facial similarities/differences and principle similarities/differences. The term facial similarity/difference corresponds to surface similarity/difference, while the term principle similarity/difference corresponds to deep structure similarity/difference.

All four combinations of facial/principle similarities/differences were created. These are labeled problem pair types A, B C, and D as defined in the Figure 5 below.

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

Figure 5. Problem pairs for the similarity rating task

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.

Figures 6a through Figure 6d below show examples of the similarity rating tasks used in the study in Interview 1.

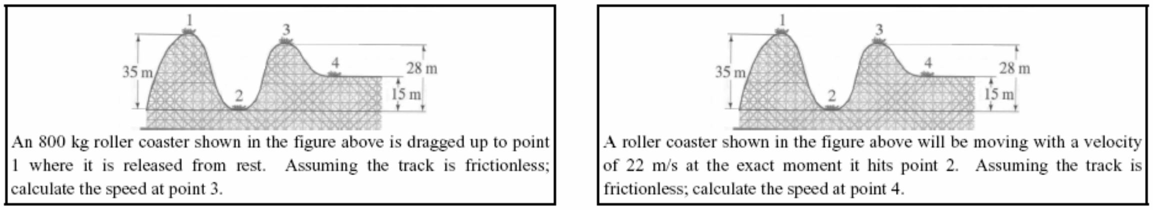


Figure 6a. Type A problem pair: Facial Similarity [FS] (both roller coasters) and Principle Similarity [PS] (both are conservative systems)

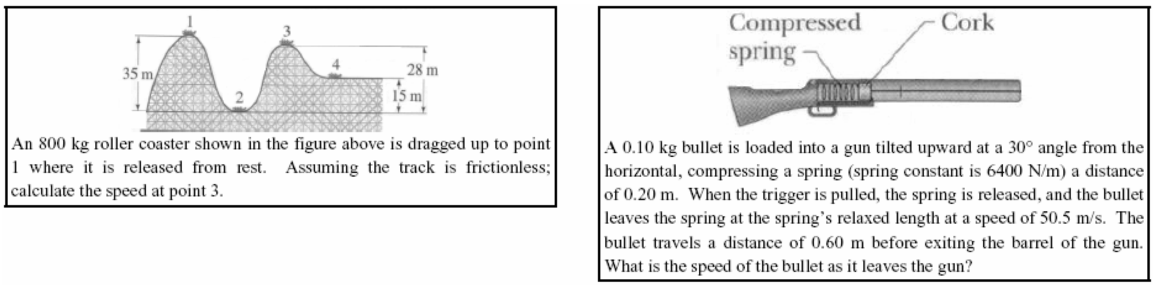


Figure 6b. Type A problem pair: Facial Difference [FD] (roller coaster vs. gun) and Principle Similarity [PS] (both are conservative systems)

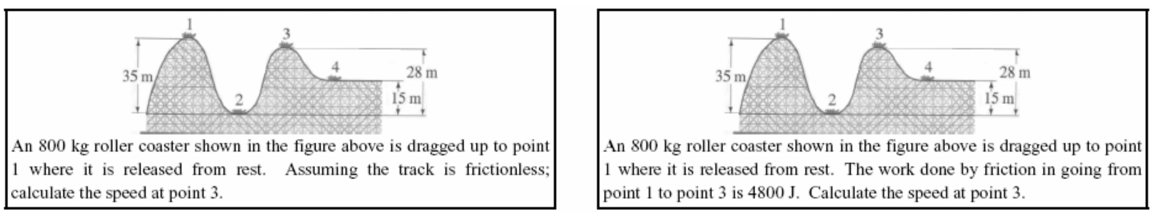


Figure 6c. Type C problem pair: Facial Similarity [FD] (both roller coaster) and Principle Difference [PD] (conservative vs. non-conservative)

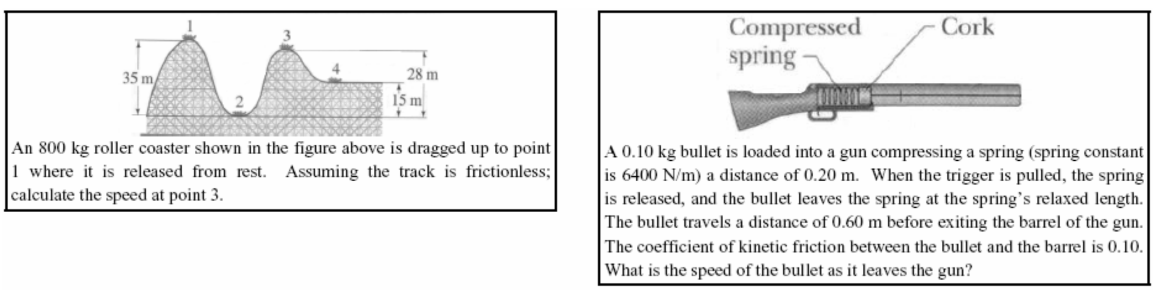


Figure 6d. Type D problem pair: Facial Difference [FD] (roller coaster vs. gun) and Principle Difference [PD] (both are conservative systems)

In addition to the problem similarity rating tasks, students were presented a sequence of problems using similar concepts for which they were asked to perform text-editing tasks (Low & Over, 1990). Text-editing tasks required students to analyze a given physics problem statement and identify whether the problem contained missing, irrelevant, or sufficient information relevant for obtaining a solution. Finally, students were presented with a challenging problem and asked to predict which of the problems they had just seen would be most and least useful as a solved example to enable them to solve the challenging problem. We do not describe the results of the text-editing and problem usefulness tasks in this paper, therefore we do not describe them the tasks here in detail.

Non-Traditional Problems on Exams

Data were also collected from five multiple choice examinations taken during the semester. Individual scores for each examination question were obtained by the primary course instructor. It is important to note the timing of the exams relative to the focus group learning interviews. Figure 7 below shows when the exams were scheduled along the timeline. As seen below exams 1 and 2 were before the protocol of our group learning interviews was finalized. Exam 3 was after the group learning protocol was mostly finalized, i.e. we had included the solved example (Problem C) into the protocol, and were providing structure in the protocol that would facilitate students to compare and contrast various problems and also reflect on the usefulness of Problem C while solving Problems A and B. Exams 4 and 5 were given after the protocol had been finalized. The difference between the finalized and 'mostly' finalized protocol is that rather than go over Problem C as we did in week 4 of the group learning interview, in group learning interviews 5 through 8 we presented Problem C to the students and asked them to go through it themselves and ask us any questions that they had about Problem C.

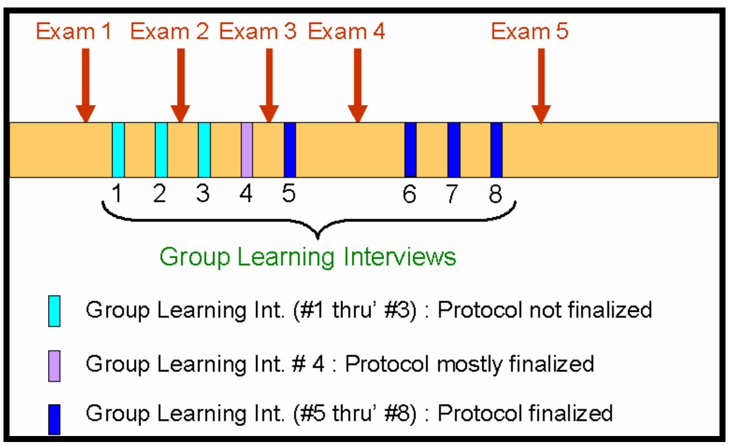


Figure 7. Timeline showing when the exams were administered during the semester relative to the group learning interviews

The last three problems on each examination were adaptations of text-editing (Low & Over, 1990), physics jeopardy (Van Heuvelen & Maloney, 1999), and problem posing tasks (Mestre, 2002). While these tasks in the original form are open-ended, the problems included on the exams were in multiple choice format for two reasons: first they conformed to the format of the rest of the test questions and second they could be graded efficiently for large numbers of students. We acknowledge that the open-ended tasks can provide richer information about the students' conceptual knowledge, but we were content with the information of students' conceptual schema provided by the multiple choice adaptations.

Text-editing tasks, as described previously, involve presenting a student with a problem statement and then asking the student to identify the missing, irrelevant, and required information in the problem statement without first solving the problem. Low and Over

(Low & Over, 1990) point out that text -editing tasks can be a measure of schematic knowledge because they require an understanding of the deep structure of the problem. Because students are asked to complete the tasks without solving the problem, students need to know the interrelationships between various physical quantities, not in terms of equations, but at a conceptual level to be able to successfully complete the task.

Figure 8 below shows an example of text-editing used on one of the class exams.

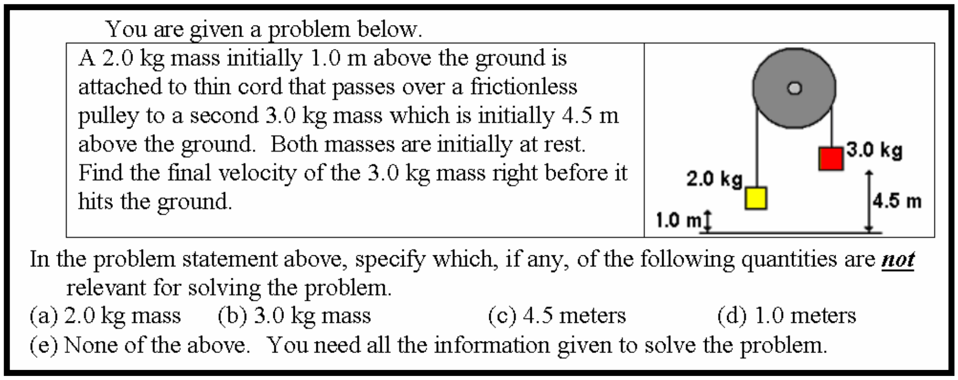


Figure 8. Example of a multiple choice adaptation of a text-editing task

Physics Jeopardy tasks were first developed by Van Heuvelen and Maloney (Van Heuvelen & Maloney, 1999). As the name indicates, these tasks require the students to work backward. Students are given a fragment of a solution to a problem and asked to identify the physical scenario that corresponds to the solution. The developers point out that these tasks require an effort to represent a physical process in a variety of ways. Because of these features, students are unable to use naïve problem solving strategies while solving Jeopardy problems.

Figure 9 below shows an example of our adaptation of a Jeopardy problem that provides students with a few steps of a projectile motion. Students are asked to determine what trajectory shown corresponds to the problem. This task requires students to relate information given in the mathematical and symbolic representation to a visual or pictorial representation.

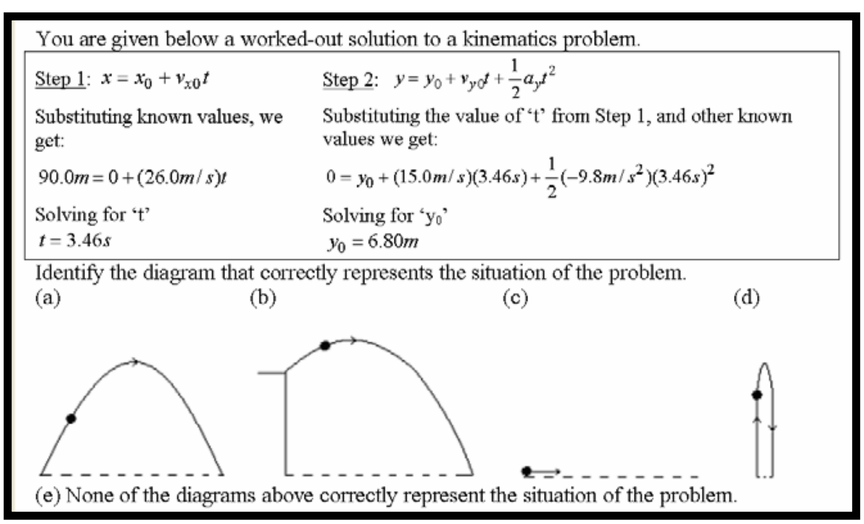


Figure 9. Example of a multiple choice adaptation of physics Jeopardy task

Problem posing tasks were used by Mestre and others (Mestre, 2002) in the context of physics problems. In the tasks presented by Mestre, students were given a scenario, typically in the form of a picture and were asked to construct a problem around the scenario that was based on certain physical principles. Mestre points out problem posing tasks are aimed at probing students' understanding of concepts as well as assessing whether they transfer their understanding to a new context. Clearly such a task was rather open-ended with multiple possible answers.

Our adaptation of this task is much more focused than Mestre's original open-ended task. It presents students with the first part of a problem statement which clearly describes a physical scenario. Students are then asked to select from a list of choices, a question, which when added to the statement will create a solvable problem that requires the use of a set of given equations. Clearly, our adaptation differs significantly from the original problem posing task designed by Mestre. First, this task clearly does have a unique correct answer. Second, it requires the knowledge of specific conceptual knowledge, represented in the form of equations. An example of our adaptation of a problem posing task is shown in Figure 10 below.

You are given the starting statement of a problem below.

A 500 kg cargo shipment, attached to a parachute, drops vertically out of a helicopter hovering 100 m above a large spring (k = 220,000 N/m). The cargo comes to rest when the spring compression is 0.50 m.

Which question, when added to the statement above, will make a solvable problem that *requires ALL of the following* equations to solve?

$$W = Fd \qquad W = \Delta KE + \Delta PE \qquad PE_{spring} = \frac{1}{2}kx^2 \qquad PE_{gravity} = mgy \qquad KE = \frac{1}{2}mv^2$$

(a) What is the speed of the cargo just before striking the spring?
(b) How much time does it takes for the cargo to make contact with the spring?
(c) What is the work done by air resistance acting on the parachute as it drops?
(d) What is the average force of air resistance acting on the parachute as it drops?
(e) None of the above.

Figure 10. Example of a multiple choice adaptation of a problem posing task

Results and Discussion

We discuss below the results of each of the sections of the study: screening interviews, focus group learning interviews, individual interviews, and performance on exams.

Screening Interviews

Most of the 21 interviewees who participated in the screening interviews had taken physics in high school. They were primarily life science majors and were currently enrolled in the only sequence of physics classes that they were required to take. This interview was conducted in the first two weeks of the semester. The views expressed by students were therefore primarily based on their experiences in high school physics classes they had taken and not necessarily based on their experiences in their current college physics class.

When asked about their study habits and use of the textbook about half of the students felt that the book was well written and easy to read. The other half felt that the book was difficult to read and limited the extent to which they would read the text.

The problem solving procedures described by students seemed to be similar to each other. They would first read the problem and pick out information that was given and asked for in the problem. Then they would solve the problem using one or a set of formulas that contained quantities that were both given and asked for in the problem. When students had difficulty solving problems, most of them reported that they would read or reread the relevant section of the textbook as well as find solved examples from the section that they believed would help them solve problems. Finally, when asked specifically about the current physics course, all of the students found the problems on the homework assignments for the first and second week to be quite easy to complete.

Based on the screening interviews, we invited 10 students to participate in the focus group learning interviews. We selected students who appeared to be interested in and looking forward to the class and were unlikely to drop out in the middle of the study. We also selected students who mentioned that they often worked with a study partner while solving homework problems.

Focus Group Learning Interviews

As described earlier in this paper, our protocol for the group learning interviews did not stabilize until week 4. Over the first three weeks we changed the protocol significantly toward providing increased procedural scaffolding to relieve students' cognitive resources to focus on reflecting about the problems rather than on simply solving them. In the last five weeks (weeks 4 through 8) of the group learning interviews we were able to provide sufficient scaffolding that would allow time for reflection.

Students were asked to reflect on and then describe the similarities and differences between the two problems A and B each given to one of the students in each pair, as comparing each of these problems with problem C, the solved example. In comparing problems students often recognize the commonality of the underlying principle. The similarities cited by students were based on the deep structure of the problems. Although they also pointed out similarities in surface features, wherever applicable, they often ranked these similarities as being less important than the similarities in the deep structure.

The differences between problems identified by students for the most part focused on the surface features. Even though some of the differences might have gone beyond surface features and might have affected the underlying mechanism of solving the problem, students seldom pointed out these differences in their comparison.

In addition to comparing various problem pairs, students were also asked to rate the usefulness of the solved example – problem C in helping them solve problems A and B. In rating the usefulness of a solved example in solving the problem, students are more likely to find the example useful if the steps in the solution of the solved example map directly onto the steps in the solution of the unsolved problem. In other words, the students look for procedural elements in the solved problem and not necessarily elements in the underlying conceptual schema to facilitate their solution of the unsolved problem. In at least one instance, students rated a solved example as being not useful because they had learned a ‘shortcut’ method for solving the unsolved problem. This was true even through students recognized that the solved and unsolved problem shared deep structure similarities. The ‘shortcut’ method was presented by the instructor in class and while it provided an efficient method for solving the problem, it did not help students think about the problem conceptually. The bottom line is that from the students’ perspective, the focus on problem solving continues to be on procedural case reuse rather than schema abstraction.

We learned that it is extremely important to design the solved example appropriately to optimize its usefulness in problem solving. The use of mathematical trickery in the solved example reduced its perceived usefulness in facilitating the solution of the unsolved problems. The difficulty level of unsolved problems must be also be carefully adjusted. If the problems are too difficult students tend focus on solving problem, not on reflection. If the problems are too easy, students do not need to reflect on what they have learned from the solved example and how it might be applicable. Finally, we also learned that it is extremely important to pose students specific questions asking them to enunciate principles of a problem and provide them with a concrete structure to facilitate reflections on similarities and differences between problems.

Individual Interviews

Students who participated in the focus group learning interviews were also interviewed individually twice in the semester – first after completing the focus group learning interview in week 5 and the second time was toward the end of the semester after completing all of the eight focus group learning interviews.

We focus on students’ ratings in the problem similarity tasks on the four types of problem pairs of type A, B, C, and D. Recall that the problem pairs contained problems that shared either or both facial similarities/differences or principle similarities/differences as described in Table 1. Examples of problem pairs are shown in Figure 6.

Before describing the students’ ratings, we should describe how we believe an ‘ideal expert’ would rate these problem similarities. Our hypothetical ‘ideal expert’ should focus exclusively on the principle similarities/differences and not at all on the facial similarities/differences between problems. Thus, this ‘ideal expert’ should rate problem pairs A (facial similarity, principle similarity) as well as pairs A (facial difference, principle similarity) as equally high on the Likert scale. This is because our ‘ideal expert’ is completely sensitive to the similarities/differences in principle and completely blind to facial similarities/differences, so although the problems in pair B are facially different, this hypothetical ‘ideal expert’ would rate the problems as being almost as similar as the problems in pair A. Based on the same reasoning, this hypothetical ‘ideal expert’ would rate pairs of type C and D equally low on the Likert scale, because they both have differences in principle, regardless of whether or not they are facially similar or different.

We now describe our students’ ratings to these four problems. We averaged the similarity ratings of each student for each problem pair type for each interview. Figure 11 below shows the rating for all four pair types for the first as well as second interview.

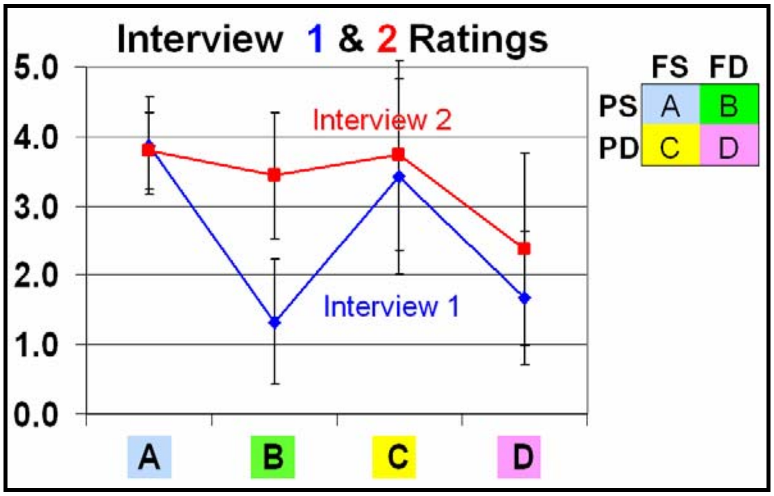


Figure 11. Students' similarity ratings of problem pairs of type A, B, C, and D for interview 1 and interview 2. The key on the top right is an abbreviated version of Table 1. It shows principle/facial similarities/differences in each type. (P=Principle, F=Facial, S=Similarity, D=Difference). The error bars are the standard deviation over all students and all problem pairs of a given type.

In 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/differences rather than similarities/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.

In interview 2, 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 which the problems have both facial and principle similarities. In other words, it appears from these data that students are able to recognize 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 statistically identical to their ratings for these pairs in interview 1. Particularly, 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. 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 are similar or different in principle.

In summary, it appears that after completing all eight weeks of the focus group learning interview, the students in our cohort group were able to discern the similarities in principle between two problems in a pair that had facial differences. But, given a pair with two problems that had facial similarities, they were unable to discern the differences in principle. This behavior appears to be consistent with the activities that they engaged in during the focus group learning interviews. Each week, problems were all focused on a single principle. We did not have problems in a given week that had any differences in principle. The only differences between the problems were facial differences. Therefore, the students appear to have developed the ability to use the facial differences as a cue to look deeper at a problem pair and decide whether there are any differences in deep structure, i.e. principle differences. This is why pairs of type B were rated highly similar in interview 2. If the problems have facial similarities, however the students appear not to look deeper to ascertain whether or not the problems have similarities in principle. The students appear to decide, based on the facial similarities, that the problems are highly similar, without attending to the underlying similarities/differences in principle. This is why pairs of type C were rated highly similar in interview 2.

An important caveat in interpreting these data should not be overlooked. In our attempt to ensure that the problems presented to students in the problem pairs were on topics that the students had covered most recently, we used the problem pairs in Figure 5 for interview 1. All of these problems were on energy conservation or the work-energy principle. Similarly, in interview 2 we used problems on the topic of simple harmonic motion that students had covered most recently in class. The differences observed in student ratings in interviews 1 and 2 could be attributed not just to the change in students' ability to discern the similarities and differences due to participation in the focus group learning interviews, but they could also be attributed to the differences in the specific problems used in each interview, the topic on which they were based, or on the fact that students were also enrolled in the class during the semester which also could have improved their abilities on these problem similarity rating tasks.

To isolate the effect of the focus group learning interviews on students' performance on the similarity rating tasks, we would need to complete interviews with students who were enrolled in the class but who did not participate in our focus group learning interviews. We would also need to have used the same set of problems for both interviews to eliminate the possibility that the effects observed are due to the specific problems being used in the interview and not the change in the students' abilities between interview 1 and interview 2.

Non-Traditional Problems on Exams

To assess students' conceptual schema in problem solving we inserted three non-traditional problems on each of the five class exams during the semester. Each exam included a text-editing, physics jeopardy, and problem posing task at the end. These problems were assigned for extra credit and presented in a multiple choice format similar to the rest of the exam.

On each exam we compared the performance of our cohort group with the rest of the class on each non-traditional problem based on a logistics test using a binomial model. We also compared the performance of our cohort group with the rest of the class as on all of the traditional problems using ANOVA single factor test. It is important to recall (Figure 7) that the first three exams were given before week 5 of the focus group learning interview during and after which the finalized protocol was used.

Figure 12 below shows the comparison of performance on traditional exam problems between our cohort group with the rest of the class on each exam, showing the mean and standard error on each exam. We find that there is no statistically significant difference (at the 0.1 level of significance) between our cohort and the rest of the class on their performance on traditional exam problems.

Exam #	Group Int. Cohort Mean ± S.E. (N)	Rest of the Class Mean ± S.E. (N)	P value*
1	75.3% ± 6.03% (N = 9)	70.0% ± 1.09% (N = 274)	0.3808
2	62.2% ± 6.11% (N = 9)	61.1% ± 1.08% (N = 274)	0.8559
3	69.7% ± 6.22% (N = 9)	65.0% ± 1.14% (N = 267)	0.4593
4	76.8% ± 4.98% (N = 9)	77.0% ± 0.93% (N = 258)	0.9795
5	79.4% ± 5.99% (N = 7)	77.6% ± 0.99% (N = 258)	0.7655

NONE are ≤ 0.10

Figure 12. Comparison of scores of cohort and rest of class on traditional exam problems

Figure 13 below shows the comparison of performance on text-editing tasks between our cohort group with the rest of the class on each exam, showing the mean and standard error on each exam. We find that there is no statistically significant difference (at the 0.1 level of significance) between our cohort and the rest of the class on their performance on text-editing problems.

Exam #	Group Int. Cohort % Correct (N)	Rest of the Class % Correct (N)	P-value*
1	44.5% (N = 9)	35.0% (N = 274)	0.5673
2	77.8% (N = 9)	74.1% (N = 274)	0.8003
3	55.6% (N = 9)	61.8% (N = 267)	0.7072
4	44.5% (N = 9)	44.6% (N = 258)	0.9339
5	42.9% (N = 7)	47.3% (N = 258)	0.3354

NONE are ≤ 0.10

Figure 13. Comparison of scores of cohort and rest of class on text-editing tasks

Figure 14 below shows the comparison of performance on physics jeopardy tasks between our cohort group with the rest of the class on each exam, showing the mean and standard error on each exam. We find that there is no statistically significant difference (at the 0.1 level of significance) between our cohort and the rest of the class on their performance on physics jeopardy tasks, except on exam 5 when the students in our cohort group performed significantly better than students in the rest of the class (p value = 0.0635)

Exam #	Group Int. Cohort % Correct (N)	Rest of the Class % Correct (N)	P-value*
1	55.6% (N = 9)	52.9% (N = 274)	0.8760
2	100% (N = 9)	92.3% (N = 274)	0.2348
3	55.6% (N = 9)	58.4% (N = 267)	0.8639
4	44.5% (N = 9)	33.7% (N = 258)	0.5127
5	100% (N = 7)	77.9% (N = 258)	0.0635

Only Exam 5 is ≤ 0.10

Figure 14. Comparison of scores of cohort and rest of class on physics jeopardy tasks

Figure 15 below shows the comparison of performance on problem posing tasks between our cohort group with the rest of the class on each exam, showing the mean and standard error on each exam. We find that there is no statistically significant difference (at the 0.1 level of significance) between our cohort and the rest of the class on their performance on problem posing tasks except on exam 4 and exam 5 when the students in our cohort group performed significantly better than students in the rest of the class (p value = 0.0012 on exam 4 and 0.0821 on exam 5 respectively).

Exam #	Group Int. Cohort % Correct (N)	Rest of the Class % Correct (N)	P-value*
1	22.3% (N = 9)	34.7% (N = 274)	0.4226
2	22.3% (N = 9)	36.1% (N = 274)	0.3741
3	11.2% (N = 9)	21.7% (N = 267)	0.4117
4	88.9% (N = 9)	36.4% (N = 258)	0.0012
5	57.2% (N = 7)	25.6% (N = 258)	0.0821

Only Exams 4 & 5 are ≤ 0.10

Figure 15. Comparison of scores of cohort and rest of class on physics jeopardy tasks

Based on the data above one can see that students in our cohort group performed significantly better than the rest of the class on two of the three non-traditional tasks (problem posing and jeopardy) on exam 4 and exam 5. The following aspects of these results are noteworthy.

First, there was no statistically significant difference between our cohort and the rest of the class on traditional problems on any of the exams. So, participating in the group learning interviews apparently did not improve the performance of our cohort group on traditional problems. These data are consistent with the notion that traditional problems are amenable to novice problem solving strategies and therefore are not effective assessment tools for gauging improvements in students' conceptual schema in problem solving.

Second, the only statistically significant differences in the data above occur in exam 4 and exam 5, which occurred after week 4 of the focus group learning interviews (Figure 6). These data are consistent with the fact that it was only after week 4 in the focus group learning interviews that we implemented the finalized protocol that explicitly required students to rank and describe the similarities and differences between problems A/B and A/C and B/C. Before week 4, students were not being provided with adequate procedural scaffolding to free up their cognitive resources to engage in reflection about the similarities and differences between the problems.

As before, an important caveat in interpreting these data should not be overlooked. The topical content of material covered on each of these exams was very different. The level of difficulty of the non-traditional problems and traditional problems on each exam was also very different. Therefore any differences between scores on traditional or non-traditional problems on exams could also be the result of these differences, rather than a result of the participation of our cohort group in the focus group learning interviews.

Conclusions

The goal of this study was to examine whether participation in appropriately designed learning activities could facilitate students' development of conceptual schema in problem solving through appropriate use of case-based reasoning. To achieve this goal we conducted a series of eight weekly focus group learning interviews with a cohort of 10 students in an algebra-based physics class over the course of a semester. Below, we address the research questions that we posed toward the beginning of this paper:

- How do students determine whether a given example is useful for solving a different problem?

In the last five weeks of the focus group learning interviews we presented students with a solved example before asking them to attempt an unsolved problem. We found that students look for procedural elements in the solved problem that map on to the unsolved problem, such as the sequence of steps followed and equations used. If students are aware of an easier procedure to solve a problem than used in the solved example, they tend to ignore the solved example completely. Even by the end of the study we found that students' focus continues to be on procedural case reuse rather than schema abstraction.

- How might we refocus students' emphasis on the similarities and differences between problems to include emphasis on deep structure differences?

We learned that it is important to choose the level of difficulty of the problems carefully to ensure that students focus not on the mechanics of solving the problem correctly, but rather reflect on the similarities and differences between various problems. The protocol of these group learning interviews evolved over the course of the semester. In the finalized protocol we provided students with a solved example as well as a structure of eliciting the underlying principles and explicitly comparing and contrasting the problems. This procedural scaffolding appeared to have enabled students to relieve cognitive resources from the task of simply solving the problem correctly and focus instead on reflecting on the problems thereby facilitating the development of deeper conceptual schema.

- To what extent does the treatment facilitate students' development of conceptual schema as assessed by non-traditional problems and similarity rating tasks?

Our results indicate that after the focus group learning interview protocol was finalized as described above, students were better able discern the principle similarities and differences between problems, only if the problems were facially different. If the problems were facially similar, it appears that students were not cued to look deeper and they appear not to have been able to discern differences in underlying principles.

Our results also indicate that participation in the focus group learning interviews does not appear to improve students' performance on traditional exam problems. But, on two of the three non-traditional exam tasks there is statistically significant difference between our cohort group and the rest of the class on exams taken after the focus group learning interview protocol was finalized.

In spite of these promising results it is important to note that in addition to the focus group interviews, students were continually studying different material in the class as the course progressed. Any improvement seen in our data could also be due to the differences in specific exam problems or individual interview tasks, as much as it could be due to the effect of the learning interventions in our focus group learning interviews.

Acknowledgements

This work was supported in part by the U.S. National Science Foundation under grant DUE-06185459.

References

Catrambone, R., & Holyoak, K. J. (1989). Overcoming Contextual Limitations on Problem-Solving Transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1147-1156.

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

Faltings, B. (1997). Case reuse by model-based interpretation. In M. L. Maher & P. Pu (Eds.), *Issues and Applications of Case-based Reasoning in Design* (pp. 39-60). Hills Dale, NJ: Lawrence Erlbaum Associates.

Hardiman, P. T., Dufresne, R., & Mestre, J. P. (1989). The relationship between problem categorization and problem solving among experts and novices. *Memory and Cognition*, *17*(5), 627-638.

Hsu, L., Brewe, E., Foster, T. M., & Harper, K. A. (2004). Resource Letter RPS-1: Research in Problem Solving. *American Journal of Physics*, *72*(9), 1147-1156.

Jonassen, D. H. (2006). Facilitating Case Reuse During Problem Solving. *Technology, Instruction, Cognition & Learning, 3*, 51-62.

Kolodner, J. L. (1997). Educational Implications of Analogy: A View From Case-Based Reasoning. *American Psychologist, 52*(1), 57-66.

Low, R., & Over, R. (1990). Text Editing of Algebraic Word Problems. *Australian Journal of Psychology, 42*(1), 63-73.

Maloney, D. P. (1993). Research on Problem Solving: Physics. In D. L. Gabel (Ed.), *Handbook of Research on Science Teaching & Learning*. New York, NY: MacMillan Publishing Company.

Mateycik, F., Jonassen, D. H., & Rebello, N. S. (2008). *Use of structure maps to facilitate problem solving in algebra-based physics.* Paper presented at the 2008 Physics Education Research Conference, Edmonton, Canada.

Mestre, J. (2002). Probing Adults' Conceptual Understanding and Transfer of Learning via Problem Posing. *Journal of Applied Developmental Psychology, 23*(1), 9-50.

Reed, S. K., & Bolstad, C. A. (1991). Use of Examples and Procedures in Problem Solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(4), 753-766.

Van Heuvelen, A., & Maloney, D. P. (1999). Playing Physics Jeopardy. *American Journal of Physics, 67*(3), 252-256.

Ward, M., & Sweller, J. (1990). Structuring Effective Worked Examples. *Cognitive Instruction, 7*, 1-39.