

Explicit Reflection in an Introductory Physics Course

Michael L. Scott, Tim Stelzer, and Gary Gladding

Department of Physics, University of Illinois at Urbana-Champaign, IL 61801

Abstract. This paper will explore a classroom implementation in which explicit reflective activities supplemented the problems students worked during class. This intervention spanned a 14 week period and was evaluated based on the relative performance between a control and treatment group. Instruments used in this study to assess performance included the Force Concept Inventory (FCI), a physics problem categorization test, and four class exams. We will discuss fully our implementation of the reflective exercises along with results from the accompanying measures.

Keywords: physics education research, explicit reflection, metacognition.

PACS: 01.40.Fk

INTRODUCTION

The field of physics education research (PER) is the scientific study of the teaching and learning of physics. There has been extensive research, both within and outside of PER, that examines the behavioral and cognitive differences between novices and experts when solving problems.[1,2,3,4,5,6,7] Research has shown that when compared to experts, novices tend to have an incoherent knowledge structure,[8,3,4] employ means-ends analysis,[1,4] tend to categorize problems by surface features,[3] and have difficulty transferring their knowledge to other contexts.[9] Experts, on the other hand, generally begin solving problems with a conceptual analysis, have a very organized knowledge structure,[3,4,8] categorize physics problems by fundamental principles,[3] and can transfer their knowledge when confronted with a new situation.

The goal of instruction is to help shift a novice to a more expert-like state in matters of knowledge, understanding, and thinking. One important behavior of expert thinking is time spent reflecting upon the meaning and structure of things learned and of tasks worked. Of interest here is how does the supplementation of physics problems with reflective exercises help facilitate the transformation of students from novices to experts. In the following research, we require students to engage in explicit reflective activities upon the completion of their worked physics problems. We then attempt to examine the effect this intervention has on students' understanding of physics.

METHOD

For this investigation, we used student subjects enrolled in the fall 2006 semester of an introductory calculus-based mechanics course at the University of Illinois, Urbana-Champaign (UIUC). The course is part of the reformed introductory physics sequence at Illinois [10] and had an enrollment of approximately 430 students. Its components included 2-75 minute lectures, a 2 hour discussion, and a 2 hour laboratory each week. This study operated within the context of that course's discussion environment, which spanned a period of 14 weeks. In this environment, students work in collaborative, small groups on a set of 4-6 physics problems, where each week the problem set focuses on one or two physical principles.[11] Nine of those weeks also included an in-class quiz which is used as a formative assessment instrument.[12]

This intervention was assessed by evaluating the relative performance of a treatment group to a control group. Four discussion classes out of twenty two were chosen for the study, with two discussion classes appointed as the treatment group and two as the control group. Both conditions had the same instructor; in addition, they also worked with identical problems and quizzes as were common to the course. However, students in the treatment group had reflective activities added to the end of most discussion problems, along with four of their quizzes including a reflective question which was substituted for a quantitative question found on the control group's quiz.

The reflective exercises accompanying each discussion problem were to engage the students in

finding meaning in the problems they were doing, were to help them see commonalities between problems in problem sets, and to assist them in knowledge integration. These exercises were placed at the bottom of each discussion problem's page and were made explicit by a prompt that said "Reflecting Back," which preceded each set. Some examples of the reflective activities include:

"Outline your strategy in solving this problem."

"Write down what you believe the intent of this problem is and/or what you learned from doing this problem."

"Compare this problem to the previous problem."

How are they similar? How are they different?"

Students were to work on these activities in their groups and were to write their responses in the space provided on their problem sheet.

ASSESSMENTS

The implementation of explicit reflective exercises as an intervention was assessed by evaluating the relative performance of a treatment condition to a control condition on set of assessment instruments. Instruments used in this study included the Force Concept Inventory (FCI), a physics problem categorization test, and four in-class exams. This section will discuss each of these instruments, how they were used to assess performance, and the results from each assessment.

Force Concept Inventory

The FCI [13] was used as an assessment in this study for two reasons. First, through reflection students should begin recognizing common principles and concepts applied over a group of physics problems that are cast in quite different contexts. We should therefore expect to see a larger gain [14] in conceptual knowledge for the treatment group than with the control group. The second reason for its inclusion is that there are numerous, well-published papers on prior findings from its use that we can draw on to compare our findings. We will thus be able to gauge our intervention against other findings.[15]

The FCI was given at the beginning of the first week of the discussion class as a pretest. As a posttest, it was given as an in-class quiz at the end of the semester. There were 33 students in the control group and 22 students in the treatment group that took both the pre- and posttest.

Prior research has shown that traditionally taught courses have an average normalized gain of 0.2 and that of courses employing more interactive engagement components have mean gains around

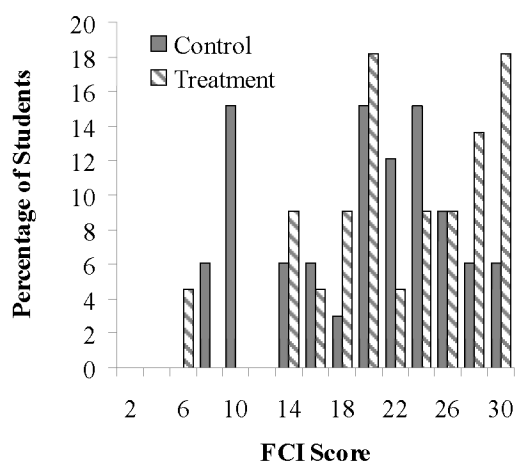


FIGURE 1. Histogram of FCI pretest scores for both groups.

0.4.[14,16] Findings from our study show that the treatment group had a normalized gain of 0.44 compared to the control group's gain of 0.41. Although the difference between the two group's normalized gains is not statistically significant, both groups had average pre- and posttest scores that were among the highest reported by universities in Hake's 1998 study. Also, the distribution of pretest scores, as shown in Fig. 1, reveals a ceiling effect. This effect limited the possible gain in scores for both groups, but this is especially true for the treatment group where one third of the group had pretest scores ≥ 28 . This high level of performance by our students limited the resolution of the FCI for measuring student gains.

Problem Categorization Test

The categorization test was based on earlier research that investigated the differences between the classification of groups of physics problems by physics experts and novices.[17,18] These studies demonstrate that experts will classify physics problems based upon the principles used to solve them (deep structure similarity); whereas, many novices will identify surface feature similarities in their classifications of problems. Reflecting on problems during a problem set, where students are forced to think about the similarities between problems, should help students to categorize problems more along the spectrum of principles than on surface features. In the Hardiman et al study, which used a similar class of students, they reported that 40% of their subjects classified problems primarily by surface features. Given this percentage, the categorization test was included to see if students in the treatment group would categorize more like experts than novices.

Modeling the Hardiman study, the categorization test was a booklet with eight pages of problems. On each page there was a Model problem that appeared at the top of the page. Beneath it were two subsequent problems (Problems A and B). Students were asked if each of the subsequent problems would be solved similarly to the Model problem, and they were to provide written explanations for their answers. There were four distinct Model problems, each appearing twice. Among each of the four comparisons per Model problem, one sub-problem had only surface feature similarity (S), one had both surface feature and deep structure similarities (SD), one had only deep structure similarity (D), and one had neither surface feature nor deep structure similarities (N).

The test was administered during the last week of class and was voluntary. To encourage serious participation, students were offered extra credit. There were 27 students in the control group and 22 students in the treatment group who completed this test.

The test was scored two ways: 1) by looking at whether students said “yes” or “no” when asked if a pair of problems would be solved similarly, and 2) by coding their written explanations. Data for the correct responses of the yes or no questions is shown in Table 1. Among the four comparison types and on overall scores, there was no statistical difference between the control or treatment groups. Both groups’ scores are slightly higher, but comparable, to the “novice” group in the Hardiman study.[19]

TABLE 1. Percentage of Correct Responses for the Control and Treatment Groups on the Four Comparison Types.

Comparison Type	Control Group	Treatment Group
S	32	25
D	60	66
SD	84	86
N	96	98
All	68	69

Note—S = surface features, D = deep structure, SD = both surface features and deep structure, N = neither surface features nor deep structure.

Student explanations were coded based whether the explanation focused on surface features, principles, equations, or physics terminology. These codes were not mutually exclusive per explanation. Subjects were then classified into three groups: (1) *surface-feature*, (2) *principle*, or (3) *mixed*. Classification into the surface-feature or principle group depended upon the subjects having those respective codes on half or more of their explanations. Pictured in Fig. 2 are the relative percentages of student types in both conditions. Juxtaposed against our data is data from Experiment 2 from the Hardiman study.[19] Both

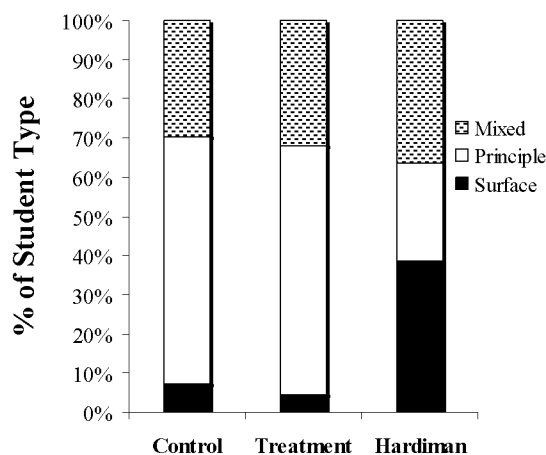


FIGURE 2. Percentages of the classification of subjects from the categorization test.

control and treatment conditions had statistically equivalent percentages of students in all three groups. An unexpected result was the large number of students in both conditions (63%) who were categorizing primarily by principle-based arguments compared to only 25% of subjects in the Hardiman study. Also surprising was that only 2 subjects in the control and 1 subject in the treatment conditions categorized primarily by surface features. Because the control subjects’ performance exceeded our expectations, this categorization test was not sufficient at discriminating between the two conditions.

Class Exams

The students were also administered four multiple-choice exams throughout the semester as part of their regular course requirements: three midterms and a final. These exams are used in the course as summative assessments of the students’ mastery of the material taught. Each midterm exam consisted of 25 questions, with the final exam having twice that number of questions. Question type varies from conceptual to quantitative, and from 3-choice to 5-choice option questions.[20]

Looking at the mean score on common questions between the two groups, the treatment group outperformed the control group at the 1.7 sigma level. Dividing those questions into conceptual and quantitative, the difference is greater for the conceptual questions with the treatment outperforming the control at the 2.3 sigma level; whereas, on the quantitative questions, the treatment group outperformed the control group at the 1.4 sigma level.

To interpret this result, we need to worry about systematic effects due to the selection of the control

and treatment groups. In our study the treatment group had a higher FCI pretest average; therefore, we used the control group's FCI pretest scores to predict how the treatment group should have performed. Looking at the difference in attained versus predicted mean score on common conceptual questions, the treatment group exceeded its predicted average at the 1.4 sigma level. Similarly, on the quantitative questions they exceeded their predicted mean at the 1.1 sigma level.

DISCUSSION

It was expected that through reflection, students would recognize concepts and physical principles across different contexts more easily than without reflection. Unfortunately, two of our tests that were intended to test this, the FCI and a categorization test, were unable to discriminate between our control and treatment conditions. This inability to discriminate may have been due to the high level of performance on both measures by both groups. When analyzing the FCI data, the treatment group had little room to improve upon its pretest scores. Analysis of the categorization test showed that most subjects in the control group are already categorizing problems by principle-based arguments. This was unexpected given findings from earlier research. The high level of performance on both these tests is probably a function of the course itself, where the entire introductory sequence at the University of Illinois was reformed to include more interactive components.[10]

To test overall physics knowledge and performance, we looked to the four multiple-choice class exams. Only on this measure was the data suggestive of an effect favoring the treatment group, and this difference was marginal at best.

Instructor observations of student interactions with the reflective activities note that many of the students appreciated the intent of the activities and worked them without complaint. There were some students, however, who saw limited value in them and would skip them until the instructor approached their table and asked about them. Another observation was that the lower ability students, who tended to sit together in groups, had difficulty just getting to the reflective exercises. These students usually had trouble with the "tools" needed to solve the discussion problems, and in general they did not complete the discussion packet in the time allotted. Whereas, students at the other end of the spectrum would finish the packet with time to spare in class. In most cases, overall exposure to the reflective activities was less than 10% of class time.

Further research on this particular study will include more extensive analyses of our measures and

exploring student responses to reflective questions found on their quizzes.

REFERENCES

1. R. Bhaskar and H. A. Simon, *Cog. Sci.* **1**, 193-215 (1977).
2. J. H. Larkin and F. Reif, *European Journal of Science Education* **1** (2), 191-203 (1979).
3. M. T. H. Chi, P. J. Feltovich, and R. Glaser, *Cog. Sci.* **5**, 121-152 (1981).
4. J. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, *Science* **208**, 1335-1342 (1980).
5. J. H. Larkin, "The role of problem representation in physics," In *Mental Models*, D. Gentner and A. L. Stevens (Eds.), Lawrence Erlbaum, NJ, 1983, 75-98.
6. Y. Anzai and T. Yokoyama, *Cog. & Inst.* **1** (4), 397-450 (1984).
7. P. T. Hardiman, R. Dufresne, and J. P. Mestre, *Memory and Cognition* **17** (5), 627-638 (1989).
8. T. deJong and M. G. M. Ferguson-Hessler, *Journal of Educational Psychology* **78** (4), 279-288 (1986).
9. J. P. Mestre, Ed. *Transfer of Learning from a Modern Multidisciplinary Perspective*, Information Age Publishing Inc., (2005).
10. C. M. Elliot, D. K. Campbell, and G. E. Gladding. "Parallel parking an aircraft carrier: Revising the calculus-based introductory physics sequence at Illinois," Forum on Education of the American Physical Society, Summer 1997.
11. The discussion problems can be accessed on the Illinois Physics Education Research group's website at <http://research.physics.uiuc.edu/PER/Materials.htm>.
12. The quizzes are free-response and consist of one conceptual question and one quantitative question with multiple parts. Students are allotted the last 20 minutes of class for the quizzes. They are graded and returned to the students the following week.
13. D. Hestenes, M. Wells and G. Swackhamer, *Phys. Teach.* **30**, 141-158 (1992).
14. R. R. Hake, *Am.J. Phys.*, **66** (1), 64-74 (1998).
15. The FCI had not been administered in this course at UIUC prior to this study.
16. E. F. Redish, *Teaching Physics with the Physics Suite*, John Wiley & Sons (2003).
17. M. T. H. Chi, P. J. Feltovich and R. Glaser, *Cognitive Science*, **5**, 121-152 (1981).
18. P. T. Hardiman, R. Dufresne and J. P. Mestre, *Memory & Cognition*, **17** (5), 627-638 (1989).
19. Comparative data comes from Experiment 2 of the Hardiman et al study. (See Ref. 7). That study involved eight Model problems, and thus had twice as many comparison types than our study. Because of the voluntary nature of our test and of time constraints, we used only four Model problems.
20. For more information on our exams, see the paper by M. Scott, T. Stelzer, and G. Gladding, *Phys. Rev. ST Phys. Educ. Res.*, **2**, 020102 (2006).