Investigating the Perceived Difficulty of Introductory Physics Problems

Elizabeth Gire & N. Sanjay Rebello

Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506-2601

Abstract. We present two studies investigating factors that correlate with students' and instructors' perceptions of problem difficulty. In the first study, introductory physics students and instructors were asked to rate the difficulty of textbook-style work-energy problems. These difficulty ratings are compared and we look for correlations between the difficulty ratings and a measure of problem complexity. We find differences between students' and instructors' ratings and a correlation between instructors' ratings and problem complexity but no significant correlation between students' ratings and problem complexity. In the second study, we asked introductory physics students and instructors to rate the difficulty of textbook-style kinematics problems. Additionally, we asked students to provide ratings of their familiarity with these problems and complete solutions. We explore the relationship between difficulty ratings, problem complexity, problem familiarity, and the rate at which students solve the problems correctly.

Keywords: Problem Solving, Difficulty, Introductory Physics, Mechanics
PACS: 01.40.Fk

INTRODUCTION

Characterizing the difficulty of physics problems is useful for both physics instruction and physics education research. Characterizations of problem difficulty can be used for constructing assignments and exams and for comparing the performance of students across courses or institutions. Such characterizations have traditionally relied on the intuition of experienced instructors and textbook authors.

Research in physics problem solving has identified several factors that influence students' problem solving performance [1]. These factors can be broadly divided into two groups. The first group includes factors that describe the interaction of the solver with the problem, like the solver's ideas about the problem domain, whether the solver has previously encountered any similar problems, and the solver's epistemological beliefs about solving problems in physics [2]. The second group includes factors that describe the intrinsic character of the problem, like how information is provided to the solver [3, 4], how well-structured the problem is [5], and the complexity of the problem's solution(s).

We explore the nature of physics problem difficulty by asking students and instructors to rate the difficulty of some textbook-style physics problems. We then compare these perceptions of difficulty and look for correlations between the difficulty ratings and the problem's complexity, students' familiarity with the problem, and the rate at which students provide a correct solution to the problem.

METHOD

We conducted two studies of problem difficulty. Both studies took place at Kansas State University, a large public university with very high research activity. The first study focused on problems in the domain of work and energy and included physics majors enrolled in the first semester of a year-long calculus-based introductory physics sequence. The second study focused on problems in the domain of kinematics and included students enrolled in the first semester of a year-long algebra-based introductory physics sequence. Instructors included physics professors and teaching assistants.

For each study, a Survey of Problem Difficulty Estimation (SPDE) was created [6]. Four versions of each survey were created and each version presented the problems in a different random order. The SPDE Work-Energy includes 16 problems and respondents were asked to rate the difficulty of each problem on a 10 point Likert scale (1=easiest, 10=most difficult). Students rated the difficulty for themselves and instructors were asked to estimate the difficulty for a calculus-based introductory physics student. Problems were either selected from a popular calculus-based textbook [7] or from University of Minnesota's Context-Rich Problems [8] or were created by the authors. The survey was administered online to both students (N=15) and instructors (N=14). Students took the survey during the week prior to an exam on work-energy.

The SPDE Kinematics includes 10 problems. For each problem, students were asked to provide a solution, rate the familiarity, and rate the difficulty. Instructors were asked only to estimate the difficulty of each problem for an algebra-based introductory physics student. Based on our experiences with the SPDE Work-Energy, we refined the rating scales by decreasing the size of the scale to a 5-point Likert scale (1=easiest or most familiar; 5=most difficult or least familiar), and we included examples of a very easy and a very difficult problem as examples of the extrema of the difficulty scale. Problems were selected from a popular algebra-based physics textbook [5] or the University of Minnesota's Context-Rich Problems or they were created by the authors. Students (N=21) took the survey in paper format and instructors (N=16) took the survey online. Students took the survey during the week prior to an exam on kinematics.

Median and mean difficulty ratings were calculated for students and instructors. The distributions of the difficulty ratings provided by students and instructors were compared using Independent Samples Mann-Whitney U Tests.

Students' solutions to the SPDE Kinematics problems were scored on a binary scale as either correct or incorrect. Solutions were considered correct if the student used an appropriate approach and only made minor arithmetic errors. All other solutions were scored as incorrect. For the purposes of analysis, correct solutions received one point and incorrect solutions received no points. This scale was selected to make the data more interpretable (i.e. the percentage of students who solved the problem correctly) and also to minimize the need for the graders to interpret students' work. Correlation coefficients were calculated for the mean student score and the mean difficulty ratings from students and instructors.

The complexity of each problem was determined using a rubric based on an Exposition-Complication-Resolution framework [10]. In the case of physics problems, the "exposition" includes sources of information about the problem situation. In textbook problems, exposition typically occurs in the problem statement. "Complications" are questions that need to be answered or quantities that need to be determined. In a well-structured problem, there is a main complication which is the overall problem to be solved. Additional complications arise during the solution of the problem and are usually thought of as steps. "Resolutions" are the answers to the complications. For example, consider the problem: "How much potential energy is stored in a spring with spring constant k = 170 N/m when it is compressed 5 cm?" The exposition identifies that a spring with a spring constant equal to 170 N/m is compressed 5 cm. The primary complication is to find the potential energy of the spring and the primary resolution is the value of the potential energy. In the solution, two additional complication/resolution pairs arise. The first is the decision of which physics ideas to use - the definition of potential energy $U=\frac{1}{2}kx^2$. The second is to match the variables in this definition with the information given in the problem statement. The complexity of the problem is determined by the number of complication/resolution (CR) pairs. This example problem would be scored as having three CR pairs.

The complexity of each problem was scored by two independent raters. The raters first scored a set of worked examples from two popular calculus-based introductory textbooks [7,11]. The inter-rater reliability for this set of problems was 0.78. The rubric was then refined. The raters used the refined rubric to score their own solutions to each of the SPDE problems and the number of CR pairs was averaged for purposes of analysis. The inter-rater reliability was 0.72 for the SPDE Work-Energy and 0.79 for the SPDE Kinematics. Correlation coefficients were calculated for the mean number of CR pairs and the mean difficulty ratings from students and instructors.

RESULTS

For both surveys, differences were found between the difficulty ratings provided by students and those of the instructors. Table 1 shows the median difficulty ratings given by students and instructors and the results of Mann-Whitney U Tests for both surveys.

For the SPDE Work-Energy, six of 16 problems received different difficulty ratings from students and instructors. Students gave one of these six problems (Problem 5) a higher difficulty rating than the instructors did. Otherwise, students rated these problems as being easier than the instructors. All of the Context-Rich Problems were rated as being easier by students than by instructors. Overall, we have discerned no pattern to explain why these problems are rated differently.

For the SPDE Kinematics, four of 10 problems received different difficulty ratings from students and instructors. Students rated all four of these problems as being easier than the instructors. None of these problems was a Context-Rich problem.

Difficulty & Student Score

For the SPDE Kinematics, the instructors' difficulty ratings correlate significantly with students' score ($R = 0.89$, $p < 0.01$) as do the students' difficulty ratings ($R = 0.76$, $p = 0.01$). The negative slope of the best fit lines indicates that as the rate of students providing correct solutions increases, the perceived difficulty of the problem decreases (Figure 1).

TABLE 1. SPDE Work-Energy Median Difficulty Ratings from Students and Instructors and Mann-Whitney U Test Comparisons. Difficulty ratings are on a 10-point Likert scale for the SPDE Work-Energy and on a 5-point Likert scale for the SPDE Kinematics.

Problem		P 01	P 02	P 03	P 04	P 05	P 06	P 07	P 08	P 09	P 10	P 11	P 12	P 13	P 14	P 15	P 16
SPDE Work-Energy	Median Student Difficulty	2	2	2	5	6	5	4	6	6	6	5	5	4	6	6	3
	Median Instructor Difficulty	2	2	2	4	5	6	6	7	4.5	6	7	7	7	5	5	6
	P (Mann-Whitney U Test)	.71	.85	.96	.10	.04	.14	.01	.24	.16	.79	.01	.01	.00	.91	.55	.00
SPDE Kin-ematics	Median Student Difficulty	1	1	3	2	2	2.5	3	2	3	3						
	Median Instructor Difficulty	1	2	3	3	3	4	4	2	3	3						
	P (Mann-Whitney U Test)	.36	.01	.07	.06	.05	.00	.01	.32	.58	.18						

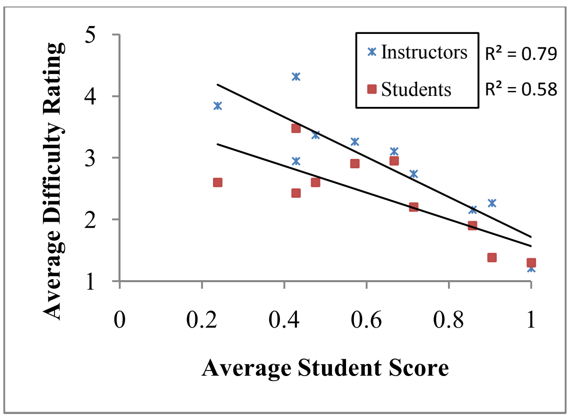

FIGURE 1. SPDE Kinematics: Difficulty Ratings for Students & Instructors vs. Average Students Score.

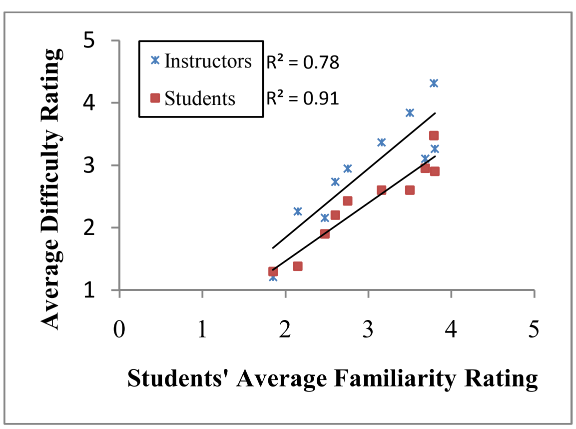

FIGURE 2. SPDE Kinematics: Difficulty Ratings for Students & Instructors vs. Students' Average Familiarity Ratings.

Difficulty & Familiarity

For the SPDE Kinematics, both students' and instructors' difficulty ratings have a significant correlation with students' ratings of problem familiarity (students: $R = 0.95$, $p < 0.01$; instructors: $R = 0.88$, $p < 0.01$). The positive slope of the best fit lines indicates that the more familiar a problem is, the less difficult it is perceived to be (Figure 2).

Difficulty & Complexity

For the SPDE Work-Energy, the difficulty ratings provided by the instructors correlate significantly with the average number of CR pairs ($R = 0.84$, $p < 0.01$). However, there is no significant correlation between the average difficulty ratings provided by the students and the average number of CR pairs ($R = 0.36$, $p = 0.20$). The positive slope of the best fit lines indicates that for instructors, as the complexity of a work-energy problem increases, their perception of the difficulty also increases (Figure 3). This relationship is not as strong for students.

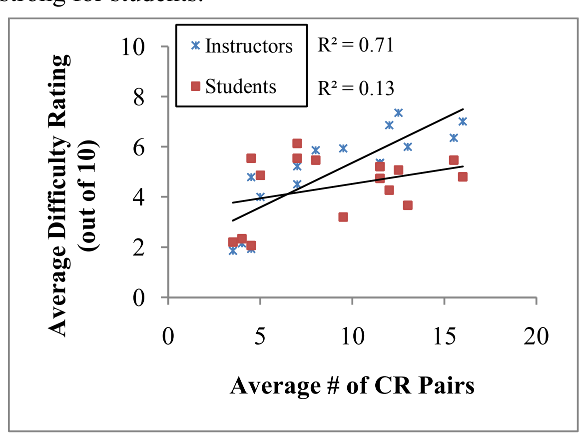

FIGURE 3. SPDE Work-Energy: Difficulty Ratings for Students & Instructors vs. Average # of CR Pairs.

For the SPDE Kinematics, the difficulty ratings provided by both students and instructors correlate significantly with the average number of CR pairs (students: $R = 0.94$, $p < 0.01$; instructors: $R = 0.91$, $p < 0.01$). The positive slope of the best fit lines indicates that as the complexity of a kinematics problem increases, its perceived difficulty also increases. The rate at which the perceived difficulty increases with complexity is similar for both the students and the instructors (Figure 4).

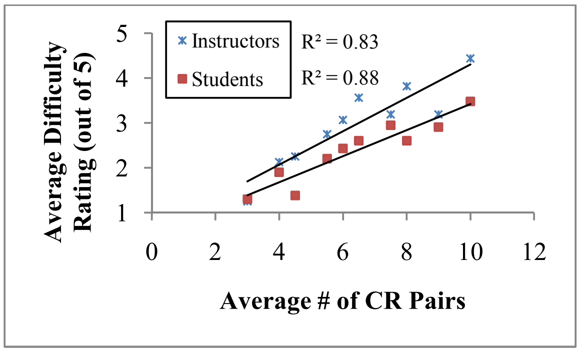

FIGURE 4. SPDE Kinematics: Difficulty Ratings for Students & Instructors vs. Average # of CR Pairs.

DISCUSSION

Students and instructors perceive the difficulty of some kinematics and work-energy problems differently, however, there is no obvious pattern that explains these differences. The interpretation of these data would benefit from interviews in which instructors and students explain the reasons for their difficulty ratings. The difficulty of both instructors' and students' ratings of difficulty correlate with students performance on kinematics problems. Interestingly, however, the correlation is stronger in the case of the instructors' ratings. One interpretation of this data may be that instructors are better able to predict the students' performance on a question than the students themselves.

Students' familiarity with a problem correlates with students' and instructors' perceptions of difficulty, and in the case of students' perceptions, this correlation is particularly strong. Indeed, asking students to rank the familiarity of kinematics problems produces nearly the same information as asking them to rank the difficulty of the problems. Familiarity is likely a very important factor in students' perception of difficulty.

In the domain of work-energy, instructors' perceptions of difficulty correlate with the complexity of the problem while students' perceptions do not. This trend does not hold in the domain of kinematics, where both students' and instructors' perceptions of difficulty correlate with complexity. A linear relationship between the number of CR pairs and the difficulty rating is not necessarily expected. For example, one could imagine that doubling the number of CR pairs might more than double the perceived difficulty of the problem. Therefore, it is somewhat surprising that a linear trend is indeed observed in these data.

These results suggest that familiarity may play a stronger role than complexity in students' perceptions of difficulty. Additionally, although the two studies were conducted with two different populations of introductory students, these results raise the question of whether students perceive problem complexity differently between the domain of kinematics (which tends to be more algorithmic) and the domain of work-energy (which tends to require more conceptual reasoning).

ACKNOWLEDGMENTS

We would like to thank the KSUPER group, especially Dehui Hu, for their continued collaboration and support. We'd also like to thank the KSU Cognitive Science Group for welcoming us to their Open Seminar. This work is supported in part by NSF Grant # 0816207.

REFERENCES

1. L. Hsu, E. Brewe, T. Foster, and K. Harper, Am. J. Phys. 72 (9), 1147 (2004)
2. D. Hammer, A. Elby, R. Scherr, and E. F. Redish, Resources, framing, and transfer (Information Age Publishing, 2004)
3. Meltzer, D. E. , Am. J. Phys., 73(5), 463-478 (2005).
4. Heller, P. & Hollabaugh, M., Am. J. Phys., 60(7), 637-644 (1992).
5. D. H. Jonassen, *Educational Technology Research and Development*, **48**(3), 63-84 (2000).
6. Copies of the SPDE Work-Energy and the SPDE Kinematics can be accessed on the web at: http://web.physics.ksu.edu/SPDE2010.
7. D. Halliday, R. Resnick & J. Walker, *Fundamentals of Physics*, 7th Edition, Wiley, 2005.
8. P. Heller, R. Keith & S. Anderson, *Am. J. Phys*, 60(7), 627-636 (1991).
9. D. C. Giancoli, *Physics: Principles with Applications*, 6th Edition, Prentice-Hall, 2004.
10. P. Baggett, J. Verbal Learning and Verbal Behavior, 18 (3), 333-356 (1979).
11. R. Knight, *Physics for Scientists and Engineers: A Strategic Approach*, Addison Wesley, 2003.