Grading student problem solutions: The challenge of sending a consistent message

Charles Henderson[a),b)]
Department of Physics, Western Michigan University, Kalamazoo, Michigan 49008-5252

Edit Yerushalmi[a)]
Department of Science Teaching, Weizmann Institute, Rehovot, Israel 76100

Vince H. Kuo
School of Physics and Astronomy, University of Minnesota,
116 Church Street, SE, Minneapolis, Minnesota 55455

Patricia Heller
Department of Curriculum and Instruction, University of Minnesota,
159 Pillsbury Drive, SE, Minneapolis, Minnesota 55455

Kenneth Heller
School of Physics and Astronomy, University of Minnesota,
116 Church Street, SE, Minneapolis, Minnesota 55455

(Received 12 February 2003; accepted 24 October 2003)

Grading sends a direct message to students about what is expected in class. However, often there is a gap between the assigned grade and the goals of the instructor. In an interview study of faculty teaching calculus-based introductory physics, we verified that this gap exists and identified three themes that appear to shape grading decisions: (1) a desire to see student reasoning, (2) a reluctance to deduct points from a student solution that might be correct, and (3) a tendency to project correct thought processes onto a student solution. When all three themes were expressed by an instructor, the resulting conflict was resolved by placing the burden of proof on either the instructor or the student. The weighting of the themes with the burden of proof criterion explains our finding that although almost all instructors reported telling students to show their reasoning in problem solutions, about half graded problem solutions in a way that would likely discourage students from showing this reasoning.
[DOI: 10.1119/1.1634963]

I. INTRODUCTION

It seems obvious that grading practices should reinforce student learning in a manner that is consistent with instructor values. No matter what directions, explanations, or examples students are given, it is the course grading practices that primarily determines what they do.[1–5] One area where an important gap has been documented between teaching practices and instructor values is in the grading of student problem solutions.[1,4] It is probable that this gap arises because of the existence of another set of hidden values that conflict with those expressed. Examining these values can help instructors understand and resolve these conflicts so that students receive a more consistent message from the class.

This paper examines the grading practices and values of a sample of 30 physics faculty from an interview based, in part, on two student solutions selected to reflect a typical conflict between instructors' stated goals and grading practices. The analysis reveals the factors that guided the grading decisions, the conflicts that arose among these factors, and the resolution of these conflicts in order to assign a grade.

The problem and two student solutions used in the interview are shown in Fig. 1. To gain a better insight into the results discussed in this paper, we suggest that the reader first assign a numerical grade to the two student solutions. The actual scale is not important because only a comparison of the scores on the two solutions for each instructor is considered in the analysis. Assume that the students who wrote these solutions were from your introductory physics class and were familiar with your testing/grading practices.

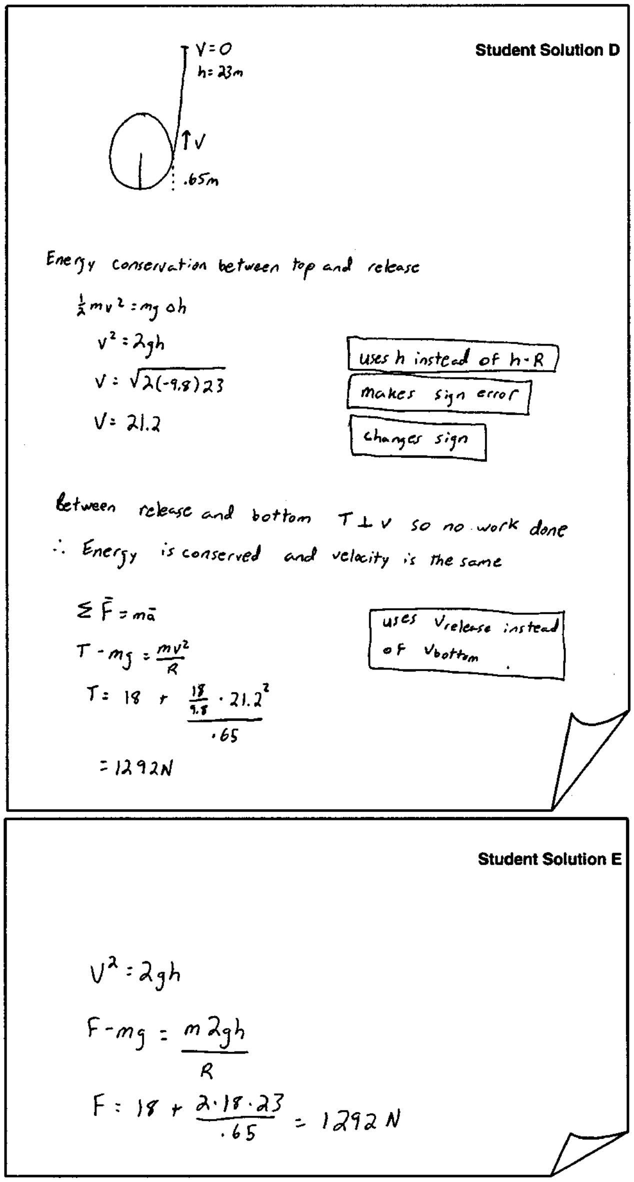
II. BACKGROUND: WHAT MOST INSTRUCTORS KNOW

We will briefly summarize four areas of general agreement about the teaching and learning of physics problem solving that provide the underlying rationale for this study.

A. Students' actual problem-solving practices

Every teacher knows that introductory physics students often solve introductory physics problems using weak problem-solving skills. Indeed, research studies document the large qualitative differences between the way experts (physicists) and novices (beginning students) solve problems.[6–10] These differences are apparent from the initial approach to a problem. Experts determine how the question relates to fundamental principles, while students often try to determine if it matches an equation from an example problem. In the process of obtaining a solution, experts plan and monitor their progress in a series of interlinked decisions while students tend to fixate on an equation or set of equations and then execute the mathematics. This approach is sometime called ‘‘plug and chug.’’ A well-known symptom of this behavior is that students often ask the instructor to solve a larger number of problems in class.

You are whirling a stone tied to the end of a string around in a vertical circle having a radius of 65 cm. You wish to whirl the stone fast enough so that when it is released at the point where the stone is moving directly upward it will rise to a maximum height of 23 meters above the lowest point in the circle. In order to do this, what force will you have to exert on the string when the stone passes through its lowest point one-quarter turn before release? Assume that by the time that you have gotten the stone going and it makes its final turn around the circle, you are holding the end of the string at a fixed position. Assume also that air resistance can be neglected. The stone weighs 18 N.



Energy conservation between top and release

$\frac{1}{2}mv^2 = mg\Delta h$

$v^2 = 2gh$

$V = \sqrt{2(-9.8)23}$

$V = 21.2$

uses h instead of h-R

makes sign error

changes sign

between release and bottom $T \perp v$ so no work done

∴ Energy is conserved and velocity is the same

$\Sigma \bar{F} = m\bar{a}$

$T - mg = \frac{mv^2}{R}$

$T = 18 + \frac{\frac{18}{9.8} \cdot 21.2^2}{.65}$

$= 1292N$

uses $V_{release}$ instead of V_{bottom}

$V^2 = 2gh$

$F - mg = \frac{m2gh}{R}$

$F = 18 + \frac{2 \cdot 18 \cdot 23}{.65} = 1292\ N$

Fig. 1. Interview problem and the two student solutions (SSD and SSE) discussed in this paper. Boxes added to the student solutions identify errors. The instructors selected for the interview were given the problem to solve before the interview. The student solutions were presented and discussed during the interview. Both students have the correct numerical answer.

B. Showing reasoning helps students learn and instructors teach

Teachers also know that self-explanations can be a powerful learning tool,[11,12] and it is not surprising that studies show that students benefit when solving a problem if they explicitly express their reasoning in a way similar to that of experts.[6,8,9,13–19] Instructors can encourage students to express expert-like reasoning by forcing students to fit their reasoning into a problem-solving framework that emphasizes the problem-solving procedures of experts. When students are consistently required to write down their reasoning within the framework of expert practice, they experience the repetition useful for learning. Reading these detailed student explanations also allows the instructor to formulate appropriate feedback.

C. Showing reasoning is an essential element of the scientific process

Of course, a hallmark of science is the prediction of real events by a logical, often formal, process of deduction from a small set of general fundamental principles and specific initial assumptions.[20] It is essential that scientific results be clearly and publicly explained so that they can be criticized and tested by others. One goal of an introductory physics course is to accustom students to this culture of scientific inquiry.[21] Problem solving can be employed as a means for students to practice accurately describing phenomena, assembling the relevant facts and principles, and logically showing how those facts and principles lead to a specific result.

D. Grading practices have a significant effect on student behavior

Every instructor hopes that students will use grading feedback to improve their skills. Indeed, several studies have shown that student behavior in a course is much more likely to be affected by grading practices than by instructor statements or other actions.[1,2,4] It is not surprising that most students choose to trust their practical experience and adjust their behavior in a manner that will get them a good grade even if they believe that alternative behavior would result in better understanding. This tendency of students to focus their behavior on getting a good grade is sometimes stated as the zeroth law of education, “if you don’t grade for it, they don’t learn it.”[22]

III. DATA COLLECTION AND ANALYSIS

We now briefly describe the interview and the background of the physics instructors who participated in the interview and provide an overview of the data analysis.

A. Interview participants

The 30 physics faculty in the sample were approximately evenly divided among four groups based on the type of institution:[23] Community College, Primarily Undergraduate Private, Research Oriented State, and Primarily Undergraduate State. The distribution of the sample is shown in Table I. The sample was randomly selected from a pool of 107 tenured or tenure-track faculty in Minnesota who had taught an introductory calculus-based physics course within the last five years and could be visited by an interviewer in a single day trip from the University of Minnesota, Twin Cities Campus.

Table I. Summary of background information for the 30 interview participants.

Type of institution	Number of instructors interviewed	Gender		Range of teaching experience	Range of teaching experience in introductory calculus-based physics	Range of typical class size for introductory calculus-based physics
		Male	Female			
Community College	7	6	1	6–35 years	3–29 times	6–75 students
Primarily Undergraduate Private	9	9	0	6–30 years	1–20 times	10–50 students
Research Oriented State	6	6	0	2–43 years	1–79 times	50–300 students
Primarily Undergraduate State	8	7	1	4–32 years	2–60 times	40–140 students

B. The interview

Each interview took about one-and-one-half hours to complete. A video camera recorded both verbal and visual responses. We focus here on the part of the interview that relates to instructor grading of student problem solutions. Prior to the interview, each instructor was given the interview problem (Fig. 1) and was asked to solve it.

During the interview, instructors were first asked to explain their purpose for grading student problem solutions. Then they were presented with five student solutions and asked to assign each solution a score (on a scale of 0–10) assuming that these solutions were from a test given in their introductory calculus-based physics course. The instructors were told to assume that the students who wrote these solutions were familiar with their testing/grading practices and expectations. The instructors were then asked to explain their grading of each solution.

The interview problem was given on a final exam in a large introductory calculus-based physics class at the University of Minnesota and the student solutions used were based on student results from that exam. Two of the solutions, student solution D (SSD) and student solution E (SSE), shown in Fig. 1, were the most useful in this phase of the interview. They were chosen to probe the potential conflict between valuing reasoning and valuing correctness. To save time during the interview, errors in the student solutions were identified to the interviewee by boxed comments as shown in Fig. 1. SSD is a detailed solution, with relatively good communication of the reasoning used in the problem-solving process. It has two mistakes that combine to yield the correct numerical answer. SSE also has the correct numerical answer, but with no explanation of the reasoning. Each of the equations that appear on SSE also appear on SSD with, at most, trivial changes. Thus, it is possible that the student who wrote SSE could have used the same reasoning as the student who wrote SSD. Of course, the student who wrote SSE might have used the correct reasoning or some other type of reasoning.

C. Data analysis

The results reported here are based on two types of data. The first is the scores assigned to the two student solutions by each of the 30 instructors. A comparison of these scores allowed us to determine if they depended significantly on the teaching environment of the instructor. One might expect scoring differences based on the amount of instructor contact with students, grading experience (most grading at the large research university is actually done by TAs), or expectations of their students. On the other hand, we might expect similar scoring because most instructors belong to the ‘‘culture of physics,’’ having had similar experiences as graduate students in obtaining their Ph.D. from a small group of large research universities.

To better understand the reasoning behind the grading decisions, we analyzed the interview statements of the six research university faculty in detail. This sample could be used because the differences among institutions were found to be smaller than instructor differences within an institution. In addition, the teaching and grading practices of this sample had been previously observed so that some of the inferences drawn from the interviews could be checked against instructor practice. In this qualitative analysis, each interview transcript was broken into statements that contained a single idea expressed by the interviewee.[24] The statements made during the grading portion of the interview were categorized and correlated to the grading decisions of the six instructors.

IV. RESULTS

A. Scoring student solutions

Figure 2 shows the scores assigned to these two solutions by the 30 instructors. The scoring of each solution differed greatly among individual instructors, especially on SSE. These individual differences were greater than those among institutions. Overall, the instructors were evenly divided between those who gave SSD a higher grade (12 instructors) and those who gave SSE a higher grade (13 instructors), with a few that graded both solutions equally (5 instructors). The interviews of the six research university instructors were then analyzed to determine the reasons behind this scoring diversity.

B. Interview analysis

An analysis of the justification of the scoring decisions of the six research university instructors resulted in three common themes: (1) Instructors say that they want to see reasoning in student solutions so they can know if a student really understands; (2) Instructors indicate a reluctance to deduct points from a student solution that might be correct as well as a duty to deduct points from a student solution that is

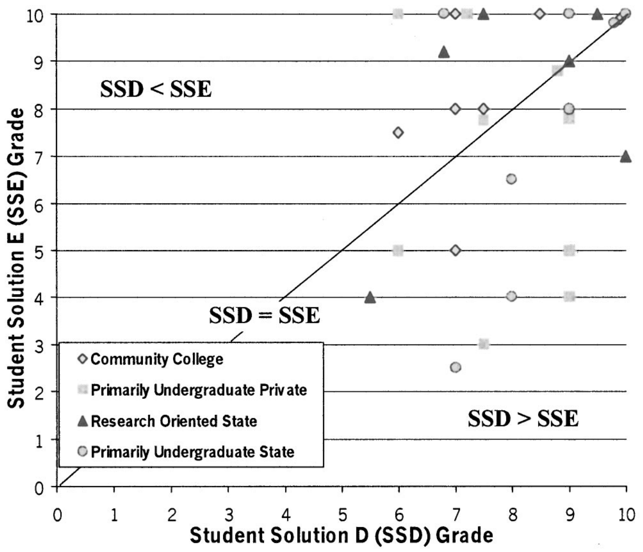


Fig. 2. Graph showing the relation between the scores assigned to the two student solutions (SSD and SSE) by all 30 of the interview participants. The straight line represents an equal score for both solutions.

clearly incorrect; (3) Instructors tend to project correct thought processes onto a student solution when the student does not explicitly describe his/her thought processes. Almost everything that the instructors said in this part of the interview reflected one of these themes.

Four of the six instructors expressed all three themes and two instructors expressed two of the three. Differences in grading appeared to result from differences in the relative strength of each of these themes between individual instructors. Although we discuss only SSD and SSE here, the three themes dominated the instructors' discourse on all five of the student solutions.

Theme 1. Five of the six instructors (except for instructor 2) indicated that seeing reasoning could help them diagnose how the student approached the problem. Theme 1 typically appeared when denigrating the lack of reasoning on SSE and applauding the existence of reasoning on SSD. An example of this theme from instructor 6 is the following, "It's hard to say whether this guy [SSE] is copying formulas out of a book or thinking, which of course is the problem with these types of solutions. I usually would tell people, counsel them away from this type of solution in a class. You know, I would say that this is like a written thing that you're trying to tell the story and you should explain what you're trying to do in applying these various things."

Theme 2. All instructors appeared to want specific evidence of student misunderstanding to justify deducting points. This manifestation of theme 2 was especially evident in the grading of SSE. Most of the instructors identified aspects of SSE that might be incorrect, but indicated a reluctance to deduct points because they could not identify a definite error. An example of this manifestation of theme 2 is given by instructor 5: "I mean, this one's [SSE] correct. There's nothing in here that's wrong. $v^2 = 2gh$ [reading from SSE]. Yeah, I mean it's not clear what v is, but of course in the end the equation would become this because at the top the velocity is zero, so you could get to that. And this one, again the student [SSE] doesn't explain where she or he got this from, but in fact you could get to this by substitution, mv^2/r, there's a 2 there, and then, yeah. So I mean this one you have to give full credit, at least I would give full credit to. It [SSE] has the right answer. It has elements of the right method. And it doesn't say anything wrong and it doesn't say anything stupid."

On the other hand, because SSD had expressed reasoning that was incorrect, most instructors used that evidence to deduct points. An example of this manifestation of theme 2 is given here by instructor 2. "The solution D made a few canceling mistakes. If I spotted those mistakes, then I would give him, probably I would end up giving him 9 points... And here this student would come to me and complain and say 'look, I got the right answer' and I would then have to explain to him that he got the right answer, it wasn't done in the wrong way, but that there were wrong things on his paper, but nothing really serious."

Theme 3. Five of the six instructors (except for instructor 1) viewed student solutions in the best possible light. These instructors were willing to believe that a student understood the physics, even in cases where the evidence for understanding was, at best, ambiguous. This tendency to assume understanding differs from theme 2 where misunderstanding cannot be proven because of lack of evidence. An example of this theme is given by instructor 2. "This guy [student E] had in mind exactly what I had in mind, namely we have conservation of energy, so that gives me the velocity, then he sees that the centripetal force is mv squared over R, so having found v squared by conservation of energy, he substitutes in, he says now it's a centripetal force problem, and bang, out comes the answer."

When assigning a score to a student solution, each of these themes can suggest a different decision. Because most of these instructors expressed all three themes, they frequently were faced with an internal conflict that they had to resolve. For example, Instructor 1 gave SSE a grade of 4. He thought that there was nothing clearly wrong with the solution and that it could be perfect (theme 2), but that there was no explanation to demonstrate that the student really knew what was going on (theme 1). He resolved the conflict in favor of theme 1, leading to a lower grade.

C. Inferences

If all themes were evident in all grading decisions, the actual score assigned must be due to the weight of each theme. A hidden construct that could determine this weight is the idea of burden of proof. We define the construct of burden of proof in the following way: A burden of proof on the instructor means that the instructor needs explicit evidence that the student used incorrect knowledge or followed incorrect procedures in order to deduct points. A burden of proof on the student means that there must be explicit evidence that the student used correct knowledge and followed correct procedures in order to get points.

For example, all three themes are evident in instructor 5's discussion of scoring SSE. He stated that the student might not be thinking correctly because little reasoning was given (theme 1). On the other hand, he stated that he believed that the student was thinking correctly (theme 3) and saw nothing incorrect for which he could deduct points (theme 2). In resolving the conflict, he placed significantly more weight on themes 2 and 3 and gave the student full credit. Thus, we infer that he placed the burden of proof on the instructor.

In a similar way, each of the other instructors was assigned a direction for burden of proof (Table II). The instructor (instructor 1) with a burden of proof on the students was very

Table II. Summary of grades given and orientation toward grading for each of the six research university instructors interviewed.

	SSD grade	SSE grade	Burden of proof
Instructor 1	5.5	4	On students
Instructor 2	9.5	10	On instructor
Instructor 3	6.8	9.2	On instructor
Instructor 4	10	7	On instructor
Instructor 5	7.5	10	On instructor
Instructor 6	9	9	On instructor

consistent in only giving credit based on explicit evidence of student understanding, that is, theme 1 was stronger than themes 2 and 3 in the scoring decisions. The other five instructors had a burden of proof on the instructor. They did not deduct points unless there was evidence that the student didn't understand or didn't follow directions. Three of these instructors (3, 4, and 6) indicated that they had grading policies that required students to ''show their reasoning.'' This policy led them to impose a small penalty on SSE, even though they believed that the student understood the physics. They hoped that this small penalty would lead the student to show more reasoning in the future. This policy differs from instructor 1, who placed the burden of proof on the students and assigned a significantly lower score to SSE than any of the other instructors.

V. SUMMARY AND IMPLICATIONS FOR INSTRUCTION

Most instructors know that there are three advantages for students to show their reasoning in problem solutions: (1) It helps students rehearse and improve both their problem-solving skills and their understanding of physics concepts; (2) It allows the instructor to observe and diagnose student difficulties; and (3) It is an essential part of doing physics. These advantages are realized only if instructors align their grading practices with their desire to see student reasoning.

The scores assigned by the instructors in this study show that less than half (40%) of them gave students an incentive for explaining their reasoning. An equal number (43%) could be viewed by students as penalizing demonstrated reasoning. An analysis of the interview data of a subset of these instructors suggests that instructors have internal conflicts when grading student solutions. Most instructors resolved these conflicts by placing the burden of proof on themselves when assigning a score to a student solution.

Many students enter an introductory physics course with neither the desire nor the skill to explain their reasoning on problem solutions. They may have learned from previous courses that showing reasoning tends to hurt their grade, or they may not have had sufficient practice to do so effectively, or both. Grading practices that place the burden of proof on the instructor, as with most of the instructors in this study, tend to encourage students to continue their initial behavior of not explaining their reasoning on problem solutions despite exhortations to the contrary from the instructor and the textbook. Students quickly learn what they perceive to be an instructor's real intent by comparing test scores with those of their classmates and can receive the unintended message that explaining reasoning is dangerous.

Our study suggests two natural strategies for an instructor to resolve internal grading conflicts in a manner that encourages students to explain their reasoning. (1) Consistently place the burden of proof on the students to demonstrate their understanding of the appropriate physics knowledge in the problem solutions they write; or (2) Consistently assign a large part of the score to the validity and quality of the student's scientific argument that justifies the result of the solution.

If the burden of proof were clearly placed on students to show their reasoning by which the answer was achieved, student solution D (SSD) would receive points for the parts of the problem done correctly, but would not receive full credit because of the incorrect physics. Student solution E (SSE), on the other hand, would receive a low score because there is no explanation of the reasoning.

If a student solution can be seen as the student's presentation of a scientific argument that supports their claim (their result), then explicit reasoning is a crucial part of the problem solution. Under this type of grading, SSD would receive points for having a relatively good argument structure, but would not receive full credit because some of the physics concepts were implemented incorrectly. SSE would receive few, if any, points because no argument structure was present.

In either of these cases, the students must be clearly informed of the criteria of evaluation, given detailed models of acceptable problem solutions for each topic that are consistent from topic to topic, and given enough time on examinations to write an acceptable solution.

ACKNOWLEDGMENTS

We wish to thank the physics faculty members who gave their valuable time to participate in this study. We appreciate the support of the University of Minnesota, School of Physics and Astronomy; the University of Minnesota, Department of Curriculum and Instruction; and the Weizmann Institute for Science, Department of Science Teaching. This work was primarily supported by the National Science Foundation (Grant No. DUE-9972470).

[a] Previously at the University of Minnesota [School of Physics and Astronomy, 116 Church St., SE, Minneapolis, MN 55455].

[b] Electronic mail: charles.henderson@wmich.edu

[1] A. H. Schoenfeld, ''When good teaching leads to bad results: The disasters of 'well-taught' mathematics courses,'' Educ. Psychologist **23** (2), 145–166 (1988).

[2] A. Elby, ''Another reason that physics students learn by rote,'' Am. J. Phys. **67** (7), S52–S57 (1999).

[3] E. Etkina, ''Formative and summative assessment in a physics class: Time to change,'' in *Proceedings of the 2002 Physics Education Research Conference*, edited by S. Franklin, K. Cummings, and J. Marx (PERC, 2002).

[4] P. Black and D. Wiliam, ''Assessment and classroom learning,'' Assessment in Education **5** (1), 7–74 (1998).

[5] G. Novak, E. Patterson, A. Gavrin, and W. Christian, *Just-in-Time Teaching: Blending Active Learning with Web Technology* (Prentice–Hall, Upper Saddle River, NJ, 1999).

[6] W. J. Leonard, R. J. Dufrense, and J. P. Mestre, ''Using qualitative problem-solving strategies to highlight the role of conceptual knowledge in solving problems,'' Am. J. Phys. **64**, 1495–1503 (1996).

[7] A. Van Heuvelen, ''Learning to think like a physicist: A review of research-based instructional strategies,'' Am. J. Phys. **59**, 891–897 (1991).

[8] F. Reif and L. Scott, ''Teaching scientific thinking skills: Students and computers coaching each other,'' Am. J. Phys. **67**, 819–831 (1999).

[9] D. Maloney, ''Research on problem solving: Physics,'' in *Handbook of Research on Science Teaching and Learning*, edited by D. Gabel (MacMillan, New York, 1994).

[10] M. Chi, P. Feltovich, and R. Glaser, ''Categorizations and representations of physics problems by experts and novices,'' Cogn. Sci. **5**, 121–152 (1981).

[11] M. Chi, N. de Leeuw, M. Chiu, and C. LaVancher, ''Eliciting self-explanations improves understanding,'' Cogn. Sci. **18**, 439–477 (1994).

[12] M. Chi and K. VanLehn, ''The content of physics self-explanations,'' J. Learning Sci. **1** (1), 69–105 (1991).

[13] A. Van Heuvelen, ''Overview, case study physics,'' Am. J. Phys. **59**, 898–907 (1991).

[14] J. Mestre, R. Dufrense, W. Gerace, and P. Hardiman, ''Promoting skilled problem-solving behavior among beginning physics students,'' J. Res. Sci. Teach. **30**, 303–317 (1993).

[15] F. Reif, ''Millikan lecture 1994: Understanding and teaching important scientific thought processes,'' Am. J. Phys. **63**, 17–32 (1995).

[16] D. Wright and C. Williams, ''A wise strategy for introductory physics,'' Phys. Teach. **24** (4), 211–216 (1986).

[17] P. Heller and M. Hollabaugh, ''Teaching problem solving through cooperative grouping. 2. Designing problems and structuring groups,'' Am. J. Phys. **60**, 637–645 (1992).

[18] F. Reif, J. Larkin, and G. Brackett, ''Teaching general learning and problem-solving skills,'' Am. J. Phys. **44**, 212–217 (1976).

[19] J. Larkin, ''Processing information for effective problem solving,'' Eng. Educ. **70**, 285–288 (1979).

[20] C. Hempel, *Philosophy of Natural Science* (Prentice–Hall, Englewood Cliffs, NJ, 1966).

[21] R. Millar, ''Students' understanding of the procedures of scientific enquiry,'' in *Connecting Research in Physics Education with Teacher Education*, edited by A. Tiberghien, E. Jossem, and J. Barojas (International Commission on Physics Education, 1998).

[22] P. Heller and K. Heller, *Cooperative Group Problem Solving in Physics* (University of Minnesota preprint, 1999). To be published by Brooks-Cole in 2003.

[23] Institutions included in each category. Community Colleges: Bethany Lutheran College, Inver Hills Community College, Minneapolis Community and Technical College, Normandale Community College, North Hennepin Community College, and Rochester Community College. Primarily Undergraduate Private: Augsburg College, Concordia University, Gustavus Adlophus College, Hamline University, Macalester College, St. John's University, St. Olaf College, and University of St. Thomas. Research Oriented State: University of Minnesota, Twin Cities. Primarily Undergraduate State: Minnesota State University, Mankato; Minnesota State University, St. Cloud; and University of Minnesota, Duluth.

[24] R. Hycner, ''Some guidelines for the phenomenological analysis of interview data,'' Human Studies **8**, 279–303 (1985).