The Role of Evaluative Abilities in Physics Learning

Aaron R. Warren

Dept. of Physics & Astronomy, Rutgers University, Piscataway, NJ

Abstract. This paper outlines ongoing research regarding the role evaluative abilities play in student learning. Evaluative abilities refer to a student's capacity for employing several general strategies for checking, critiquing, and judging information. In the context of physics, such information may include proposed problem solutions, proposed models, and conceptual claims. Strategies for evaluating information include dimensional analysis, checking special and limiting cases, and analyzing the assumptions which underlie a model, solution, or claim. Over the past year I have been developing and testing a number of activities designed to promote and assess students' evaluative abilities. I report preliminary results regarding the correlations between students' evaluative abilities and their understanding of physics, and identify future avenues of research.

MOTIVATION & BACKGROUND

Reports from the AAAS[1] and NSF[2] emphasize that science courses ought to cultivate a wide array of abilities. In particular, these reports agree that students should develop those abilities which allow them to become self-corrective learners. The ability to critically evaluate information is a central component of self-corrective learning. Evaluative abilities allow students to identify mistakes in their thought and work, a necessary first step towards correcting such mistakes.

By the term "evaluative abilities," I mean the ability of a student to use any of several general strategies for checking and judging information[3]. There are several types of information which a student typically produces. Students often propose problem solutions (e.g., while working out homework problems), which entail the use of a model to calculate or predict some quantity. Students may also work to answer conceptual questions in which they must make some claim by employing their qualitative understanding of physics. Finally, students may propose models for specific situations in order to explain or predict the outcome of an experiment.

I focus on a few general strategies for evaluating these types of information. Dimensional analysis is a powerful technique for determining whether a given equation is physically self-consistent. Checking special and limiting cases determines how robust a certain model may be by testing whether it is consistent with other models or conceptual ideas. Similarly, identifying and checking the assumptions which underlie a model allow one to determine how accurately a model can explain or predict a system's behavior in a given situation.

Theoretical Perspective

There are several arguments for why enhancing students' evaluative abilities should enhance student problem-solving performance (on both conceptual and traditional problems), and each argument appeals to a different theoretical perspective of knowledge and learning. One such perspective relies on the notion of schemas. While there are other theoretical perspectives which may be used (e.g., based on neuroscience, or conceptual change theory), I leave such discussions for future publications.

A schema is a pattern of activated knowledge incorporating facts, rules, p-prims, facets, and other knowledge elements. Research suggests that many of the differences between students and physics experts are due to differences in the content and organization of their schemas[4]. Experts have schemas which incorporate knowledge into a hierarchical, globally coherent pattern. Students learn by acquiring new knowledge elements, and by developing or changing links between knowledge elements. In this context, the use of an evaluative strategy may be represented as a schema (which I call a *meta-schema*) that is linked to an array of schemas responsible for processes such as problem solving. When a particular schema is activated, an evaluation meta-schema can be triggered. The use of an evaluation meta-schema can lead to the restructuring of the triggering schema, or of links attached to the triggering schema, by causing the student to recognize incoherencies among schema and inconsistencies between the product of the triggering schema (e.g., a proposed problem solution) and a desired result (e.g., a correct problem solution). Evaluation provides feedback which determines whether a schema is inappropriately activated, and whether the schema is sufficient for some task.

Developing & Assessing Evaluative Abilities

Although evaluation abilities are likely to play a significant role in learning[3], they receive relatively little attention in most curricula. Over the past year, I have been developing a set of activities which are designed to develop and assess evaluative abilities among physics students (available for download from http://paer.rutgers.edu/ScientificAbilities/FormativeAssessmentTasks/default.aspx). To maximize the effectiveness of the evaluation activities, I have structured them as formative assessments. Formative assessment activities appear to be among the most powerful instructional intervention methods currently known[5].

Formative assessment tasks generally follow three guiding principles[6]. First, a task must inform students of the learning and performance goals. Secondly, a task should inform each student of their current level of understanding. Finally, a task should help guide each student to achieve the performance goals by developing strategies and skills. The general role of formative assessment, then, is to supply real-time feedback to the students and instructor which identifies a goal state, each student's current state, and a method for each student to progress from her current state to the goal state.

In order to assess and develop students' evaluative abilities, they must first be given something to evaluate. Therefore, each evaluation task will present some proposed problem solution, proposed model, conceptual claim, or experiment report which the student must critically judge. To satisfy the first requirement for formative assessments, I have developed rubrics, guidelines, and examples which inform the students of the nature of the tasks and how their responses will be graded. The rubrics identify four classes of possible responses, with each class indicating a certain level of ability development. Guidelines and examples inform the students of strategies to be used when responding to an evaluation task. By discussing these guidelines and examples with their peers and instructor, students can begin to develop an understanding of each evaluative strategy.

To satisfy the latter two conditions for formative assessment, the students are presented with many opportunities to develop their evaluative abilities. In recitations, students work on evaluative tasks and discuss them with each other and with instructors. In laboratory, students work on evaluative tasks as part of an exit interview process where each student group discusses the tasks among themselves and with the instructor. In lecture, students discuss evaluative tasks in small groups. These activities allow the students to understand how their current level of performance compares with the goal performance, as well as how to improve their performance.

Research Goals

We report on research designed to conduct an initial investigation of the role evaluation plays in student learning and problem-solving performance. This research constitutes a trial study to determine whether more extensive investigation into this idea may be worthwhile. Based on the theoretical discussion above, I expect strong evaluative abilities to correlate with enhanced student learning and better problem-solving performance. As an initial test of this claim, I can measure students' evaluative abilities (using the activities I developed) and determine whether there is a significant positive correlation with student scores on exams.

One may suspect on theoretical grounds that the development of evaluative abilities actually causes students to develop a robust understanding of physics. Although a significant positive correlation would not imply any causal link, it would be consistent with such a hypothesis and therefore indicate that further investigation of this issue is worthwhile.

The validity of any measured correlation relies upon two major assumptions. First, I assume that our activities do indeed measure students' evaluative abilities. Second, I assume that exam scores accurately reflect each student's understanding of physics, at least to some degree. To test both of these assumptions I can conduct student interviews to gather more detailed data on student understanding and compare these data with each student's score on evaluative activities and exams. Such interviews will be conducted during the upcoming academic year.

RESEARCH

Data Collection & Coding

During the 2003-4 academic year evaluation activities were developed and administered to students in an algebra-based introductory physics course for life science majors at Rutgers University. The course had an enrollment of roughly 150 students. Most of the activities were administered during the weekly laboratory sessions. Here, students worked with their lab partners (3-5 students per group) to complete several evaluation activities. The evaluation activities were designed around the subject matter which had been dealt with in that week's lectures, workshops, and labs. When a student group felt it had completed the evaluation activities, they presented and discussed their work with other lab groups and with the lab instructor. These discussions gave students the opportunity to make repeated attempts at understanding and achieving the goals of the activities while incorporating feedback from their peers and instructor, and constituted the formative assessment implementation of my evaluation activities.

As part of this course, students had to take two lab practicals each semester. The lab practicals included short-answer format questions which required a written response. Student responses were photocopied before being graded by the lab instructors, who used the rubrics I developed in order to assign grades. Thus, the lab practicals served as a summative assessment tool.

The photocopied student responses to the first and last practicals of the year (practicals 1 and 4) were used as our raw data. Each evaluation activity on the lab practicals was designed to evoke the use of a particular evaluative strategy. Therefore, student responses to each question were coded in order to measure the level of performance with a specific evaluative strategy. The coding assigned an integer between 0 and 3 to each student response. Similarly, students' answers to the conceptual questions were coded on an ordinal scale from 0-3 using a taxonomy developed by the course teaching assistants.

There were six different topics covered on each practical. Each student was randomly assigned a topic at the practical. Each practical had at least two evaluation questions, and each question required the use of a different evaluative strategy. This design gave an average of 65 students on each practical (this number varied between 64 and 72) for which we could determine the correlation between a specific evaluation ability and performance on corresponding conceptual questions from the practical, as well as the correlation with performance on the multiple-choice exam. This total of ~65 students for each correlation was obtained by combining student scores for a specific evaluation activity regardless of the physical context of the activity. This was done in order to have a sufficient number of data points for calculating correlations.

It is not clear whether such a combination of scores from different contexts should affect the correlations. The possible context-dependence of evaluation abilities is something we will address in future work. While the actual implementation of an evaluation strategy certainly depends on the topic at hand (e.g., needing to know the dimensions of a certain quantity), it may be that the actual correlations between evaluation abilities and conceptual understanding/exam performance are independent of context. Investigation of this matter is necessary to test the robustness of our current results. I am currently repeating this study with more subjects in order to address this issue.

Preliminary Results & Discussion

My coding for each practical was subjected to inter-rater reliability checks with three other members of the Rutgers Physics Department. The overall agreement level was 96%, and the agreement level for each rubric varied between 91% and 99%. To examine the relationship between student evaluation abilities and student performance on exams, I calculated Kendall's Tau-b. Those correlations which were significant (at least at the 0.05 level, 2-tailed) for both the first and fourth practicals are shown in Table 1.

I also note that the correlation between conceptual scores and exam scores was consistently significant (Practical 1, $\tau_b = 0.293$, $p = 0.000$, N=135; Practical 4, $\tau_b = 0.272$, $p = 0.000$, N=169). This supports our assumption that student exam scores reflect their understanding of physics.

TABLE 1. Significant Nonparametric Correlations

Evaluative Strategy Type	τ_b (Practical 1) p (2-tailed)	τ_b (Practical 4) p (2-tailed)	Measurement of Understanding
Limit/Special Case Analysis	0.294	0.323	Conceptual Score
	.007	.000	
Limit/Special Case Analysis	0.312	0.227	Exam Score
	.002	.003	
Dimensional Analysis	0.358	0.275	Conceptual Score
	.002	.009	
Assumption Analysis	0.244	0.357	Conceptual Score
	.020	.003	
Assumption Analysis	0.292	0.292	Exam Score
	.003	.008	

The results in Table 1 indicate that evaluative abilities generally correlate with students' understanding of physics. However, dimensional analysis did not significantly correlate with exam score, although it did correlate with conceptual scores as shown. This indicates that dimensional analysis is related to students conceptual knowledge, but is largely independent of their quantitative work. On the other hand, limit/special-case analysis and assumption analysis each correlated with both conceptual and quantitative problem scores. These results seem to suggest that different evaluation abilities correspond to different aspects of students' knowledge and use of their knowledge.

I therefore feel that the structure, function, and evolution of evaluation abilities warrant more thorough study. In particular, I would like to determine whether there is in fact a causal connection between evaluation abilities and learning. Although the current results are consistent with the hypothesis that evaluation abilities are necessary for robust learning, they of course do not warrant such a conclusion because correlation does not imply causation. An alternative explanation may be that there are some confounding variables (e.g., effort, interest, prior knowledge) which are responsible for the measured correlations. I will be conducting research, using both large-N and small-N techniques, to further study the relationship between evaluation abilities and learning.

With the help of Alan Van Heuvelen and Sahana Murthy, I will conduct a large comparison group study between two parallel introductory large-enrollment classes to test whether improving evaluation abilities causes improved exam performance. One course (experimental group) will include my activities as part of recitations, homeworks, and labs, while the other course (control group) will not. Both courses will include evaluation activities on the exams to measure evaluation abilities, and several multiple-choice problems on the same topic. We expect that only students from the experimental group will significantly improve their evaluation scores over the duration of the course (indicating my activities aid the development of evaluation ability). We also expect that as the experimental group's evaluation scores improve, they will significantly outperform the control group on the multiple-choice exam problems (indicating that improved evaluation abilities caused improved exam scores). Videotaped student interviews will be used to triangulate the data.

ACKNOWLEDGEMENTS

I would like to thank Alan Van Heuvelen, Eugenia Etkina, Marina Milner-Bolotin, David Brookes, Michael Gentile, Sahana Murthy, and Jessica Warren for help with the implementation of my tasks, inter-rater checks, and for feedback on this manuscript. I also thank the anonymous referees for valuable feedback. Supported in part by NSF grants DUE 0241078, DUE 0336713.

REFERENCES

1. American Association for the Advancement of Science, Project 2061, *Benchmarks for science literacy*. New York: Oxford University, 1993.
2. NSF Directorate for Education and Human Resources Review of Undergraduate Education, *Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology*, 1996. Recommendations may be seen at http://www.ehr.nsf.gov/egr/due/documents/review/96139/four.htm
3. Anderson, L.W. & Krathwohl, D.R., *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. Longman, 2001.
4. Sabella, M.S., Ph.D. Dissertation. University of Maryland, 1999.
5. Black, P. & Wiliam, D., *Inside the black box: Raising standards through classroom assessment.* London: King's College, 1998.
6. Sadler, R., *Instructional Science,* **18**, 119-144 (1989).