On the dissemination of proven curriculum materials: RealTime Physics and Interactive Lecture Demonstrations

Michael C. Wittmann

Department of Physics and Astronomy and the College of Education and Human Development at The University of Maine

5709 Bennett Hall, University of Maine, Orono ME 04469–5709

wittmann@umit.maine.edu

tel: 207 – 581 – 1237

0140Gm, 0140Fk, 0150Qb

Abstract

RealTime Physics (RTP) and Interactive Lecture Demonstrations (ILD) have had great success in improving student learning at their development sites, but dissemination of research-based curriculum materials often shows that student learning at secondary implementation sites is less than at primary, development sites. Five schools using RTP and ILD materials were investigated during the first three years of the implementation of the materials. Student learning of the relevant physics was investigated through the use of several standardized tests (including the Maryland Physics Expectations Survey, the Force and Motion Concept Evaluation, and the Electric Circuits Concept Evaluation). Results show that normalized gains of pre- and post-test scores are larger for classes effectively implementing the RTP and ILD materials than for classes using the regular

instruction methods or having serious implementation problems. Furthermore, post-test scores are not as high as reported by the development sites.

Introduction

Ongoing research in physics education[1] has shown that students often do not learn the physics adequately in a traditionally taught classroom, while evidence has shown[2-8] that modifying the curriculum based on research into student learning can have a profound and positive effect on student understanding of the material. Much of the research and curriculum development have been carried out at the same institution, creating a good match between the needs of the students and the curriculum materials themselves. As these materials are used at other schools, new issues may arise. There may be a mismatch between the curriculum and students at these institutions, or there may be a mismatch between the curriculum and instructors who are unfamiliar with the specific implementation guidelines, strengths, and shortcomings of the materials.

To understand how well the RealTime Physics and Interactive Lecture Demonstration dissemination project (referred to as the RTP/ILD project in the rest of this document) helped students learn the physics in their classes, we must consider two facts. Most importantly, student understanding of the physics must be evaluated. Direct studies of the students themselves give the most complete picture of how effective the curriculum is. Second, the quality of curriculum implementation at the different test sites must be reported. Success in implementation directly affects students in the classroom. Furthermore, issues arising at each university may help future users more effectively implement the materials when they choose to use them.

In keeping with the goal of focusing on student learning of the physics, this paper only describes how implementation affected student learning, and does not include administrative issues that are outside the scope of a study of student learning in the modified classroom. I first describe the tools used to measure student learning and test site implementation. I then describe results from the investigations into student understanding and attitudes toward the course. Finally, I describe how different schools had difficulties in adequately using the RTP/ILD materials and what effect this had on student learning.

I wish to state at the outset that this work was aided by Chris van Breen (who acted as assistant in this study during the middle part of the evaluation period) and Hadley Lawler (who guided the study in its first year) at the University of Maryland. Also, Edward F. "Joe" Redish and the Physics Education Research Group at the University of Maryland provided support during the course of this project.

I. Evaluation tools

Four conceptual tests and one attitudinal survey were used to gather information about student performance at the implementation sites. For reasons explained below, I will only report in detail on the results from two of the conceptual tests. In addition, instructor comments were gathered through the use of surveys at different points of the implementation. In this section, each of these data sets is described in more detail.

A. Investigations of Student Performance

Standardized conceptual tests and an attitudinal survey that measures student expectations in the course were used at the test sites to evaluate changes in student understanding of the material and attitudes toward physics learning. These tools were chosen because they gave the most complete information about the overall success of the implementation. Results from before instruction (pre-testing) and after instruction (post-testing) were studied, as is commonly done in physics education research.[1]

Conceptual Tests

Research has shown that students have deep conceptual difficulties with the material they learn in the physics classroom.[1] The primary goal of the RTP/ILD classroom was to improve student learning of physics. To assess this goal, we used several previously developed conceptual tests.[8] Due to logistical limitations, no interviews were carried out with students at the test sites. This decision limited the type of data we could get from test sites, though previous work[9] has shown that the concept tests alone provide meaningful data about the quality of student learning in a modified setting.

`The four tests used in the study are the

- Force Concept Inventory (FCI),
- Force Motion Conceptual Evaluation (FMCE),
- Electric Circuits Conceptual Evaluation (ECCE), and the
- Heat and Temperature Conceptual Evaluation (HTCE)

In this paper, results from the FMCE and ECCE will be described. The tests have been developed through a long process of research into student difficulties. Presently, a large and growing database of student results exists for each of the tests.[8-10]

Attitudinal and expectations tests

A second type of survey test was used to gather data about student expectations of the physics classroom. The study of expectations allows insight into students' attitudes about the hidden curriculum, the implicit expert description of the community of physics. We use the term expectations to describe the attitudes, beliefs about the classroom, and epistemologies that students bring to the classroom. Research has shown that students approach the classroom with a set of expectations that are unfavorable to properly learning physics in a qualitative, coherent manner that is linked to the real world around them.[11,12]

The Maryland Physics Expectations Survey (MPEX) investigates student expectations through the use of a 35 question, Likert-scale survey. Students are asked to agree or disagree with a set of statements that give us insight into how they view the physics classroom. Clusters of questions have been designed to view student expectations on a variety of topics: coherence of physics, conceptual understanding of physics, student independence in learning physics, the link between physics and the real world, the link between physics and mathematical formalism, and the effort needed to learn physics.[9,12] Though questions have been raised about what attitudinal tests such as the MPEX actually measure,[13] these tests are valuable in that they focus an instructor's attention on student productive or unproductive expectations for learning in the classroom.

Gathering Student Data

We used only matched data (e.g. students had to submit responses both before and after instruction to be counted in the study). Not analyzing other students prevents us from using pre-instruction data from those who dropped the course even as it restricts the use of post-instruction data from those who, for whatever reason, did not answer the survey at the beginning of the semester. Both the conceptual and the expectations surveys have been asked in this fashion.

The conceptual and the expectations survey data were collected in two different fashions. Students took the conceptual tests before and after instruction, filling out answers for scanning machines during either lecture or laboratory time. Student responses were sent to the University of Maryland where they were scored and evaluated.

Students answered the MPEX survey questions both in the fashion of the conceptual tests and on-line during out-of-class time. The Project Galileo team at Harvard University[14] placed the MPEX survey on-line in a form accessible on the World Wide Web. Students logged in on their own time and took the survey within a several day long time window specified by the instructor. Scores were compiled by the Project Galileo team and then sent to the evaluation site at the University of Maryland.

A slight modification of the outside-class administration of the MPEX was used at two schools in the dissemination project. Students were given the scantron sheet (as used with in-class testing) and the actual test to take home. The forms and surveys were due in the next class. As with the on-line administration of the test, students were free to take it at any time they like.

Both forms of out-of-class testing were introduced due to limitations of the original MPEX testing in the first year of RTP/ILD implementation. A short version of the MPEX survey (focusing on only a few of the analysis clusters in the test) had been developed to allow students the possibility of answering both the FMCE and the MPEX during a single class period. Nevertheless, many students did not answer the MPEX questions completely. Possibly, they did not have enough time. In addition, the expectations they described may have been strongly influenced by the priming provided by the FMCE. As a result, the interpretation of MPEX data from the first year of the RTP implementation is rather problematic and will not be included in this paper (except where noted). The non-computer administration of the MPEX outside of class was chosen in two locations where computing use caused serious implementation issues. It was a compromise based on a desire to get the MPEX data in any possible form rather than no data at all.

Control Studies

To get effective and meaningful data about the effectiveness of implementation, it would be best to have a control group of students from the same institution, taught by the same instructor, with as many other instructional variables as possible held constant with respect to the RTP/ILD implementation. Since schools usually used the materials in all their classes, this was often not possible at a test site. Fortunately, the data have shown that, in general, student performance in various classrooms is accurately described regardless of instructor.[1] As has been shown by Hake[10] and Saul,[9] interactive engagement classrooms usually provide a better setting for student learning than traditional instruction classrooms do. But, as Cummings et al.[15] have pointed out, the interactive engagement activities must be implemented correctly, otherwise large expense leads to little result. This issue will be discussed later in this report.

The control data for the RTP/ILD dissemination come from a variety of sources. Most important are the classes taught without RTP or ILD materials at the institutions themselves. Two outside sets of data also serve as controls. First are the traditional instruction classes at different institutions that have used the conceptual and expectation tests used in this study. These schools set a type of baseline for the schools in which in-school, pre-RTP/ILD testing was not done. Data from Thornton and Sokoloff's work [8] and other data gathered at the University of Maryland[9] will be used for this purpose. Other schools that have implemented RTP/ILD materials also serve as comparisons. Primary implementation sites, such as University of Oregon, usually serve as examples of the best possible results.

One set of data in this study serves dual purposes. Though in general it is a result of the study to find those schools in which there was a weak or problematic implementation of RTP materials, results from such classes serve a further purpose. They help replicate the findings of Cummings et al. with respect to the effectiveness of partial implementation of research-based curricula. A later section of this report discusses how the effectiveness of implementation is determined.

B. Level of Implementation

A variety of methods of assessing the level of implementation were used in the RTP/ILD dissemination project. Mid-project interviews and end-of-project surveys provided direct feedback on implementation issues. In addition, the implementation of evaluation tests and surveys provided direct evidence of how instructors worked in the classroom and used the materials.

Presentations of and Interviews with Team Leaders

In October, 1999, a team leader meeting was held in Oregon. Team leaders presented their implementation plans and discussed issues relevant to their schools, together with preliminary findings from the first year of the project. During this meeting, the project leader (David Sokoloff) and evaluation leader (Michael C. Wittmann) met with team leaders for individual interviews.

Surveys on Implementation and Evaluation Issues

During the Oregon meeting, 1999, team leaders answered questions designed to elicit information about implementation issues and student learning in their classes. These surveys asked team leaders to address two different sets of issues. The first was the role of the diagnostic testing in the classroom. Since testing takes time from a class that has limited class time, we were interested in seeing what benefits instructors saw in the testing. The second was to describe implementation issues at their schools.

This same survey, in slightly modified form, was given again at the end of the project.

II. Methods of analysis for evaluation of student learning

In this section, we describe the analysis of the conceptual test data and the expectations survey data. Methods from the literature (see especially [9,12]) were used to analyze the data. In each case, we also describe our predictions for the results in light of previous research.

A. Conceptual Test Data

Matched data from before and after instruction were compared. As stated above, only the FMCE and ECCE results are described in this paper. Student performance on each of the tests was evaluated in terms of correct scores, with the FMCE being broken down into "clusters" of answers that illustrated special elements of student understanding.

The central element of the analysis was the "normalized gain" of the average scores. "Average normalized gain" describes the fraction of the possible change in average score of a class. Thus, a class whose FMCE score went from 20% to 40% correct would have a normalized gain of 0.25 (having moved 20% of a possible 80%, .20/.80 =

0.25). Normalized gain factors (or, for brevity, “gains”) have become an accepted measure of diagnostic testing because they measure a class's performance against its own pre-instruction performance. Results have shown that schools with similar types of instructional styles have similar gain levels even when the pre- and post-test scores are different.[9,10] Thus, the normalized gain gives a reasonable and meaningful description of the success of a form of instruction. Additional possible methods of using the gain to evaluate student learning in the classroom[16] were not used in this analysis, though future evaluation of this sort may be possible.

To compare instruction methods, a histogram of normalized gain scores was created (for example, see Figure 1). The histogram gives the number of classes scoring in each range of normalized gains, while also indicating different types of instruction. This representation gives insight into the range within which each form of instruction can affect a class’s understanding of the material, as measured by a given standardized test.

The data analysis of the FMCE was noteworthy for two reasons. First, the data were analyzed using a rubric provided by Ron Thornton for proper analysis of the full 47 question FMCE test. The rubric calls for only 35 of the questions to be analyzed. Of the others, 4 are on energy topics that are not part of an overall understanding of force and motion, and the rest fall into two categories. First are "priming" questions designed to get students prone to thinking a certain way to actually think that way on subsequent questions. Second, some questions are easily answered correctly for incorrect reasons.[17] We did not analyze either type of question for the simple reason that they give little information about student learning during the semester. In addition, some questions are analyzed in groups, meaning that students only get full credit (weighted to two points rather than three) on a topic when they answer three questions of a set correctly. (More detail on this analysis method and the actual spreadsheet templates used to carry out the analysis can be found at URL http://perlnet.umephy.maine.edu/materials/.)

FMCE analysis was also noteworthy because it is possible to break the test responses into clusters of questions with common ideas and concepts. The FMCE addresses topics that stretch over the course of nearly a full semester of instruction. As a result, there is value in understanding student responses to individual parts of the test (such as those dealing with velocity, Second Law, and so on), and it has been designed to meet this goal. Further details will be given below, when discussing student data.

The ECCE is a much shorter test, since it focuses on material that typically is taught for a shorter time than is force and motion in a typical introductory course. Thus, it was not analyzed using clusters of responses.

The HTCE and FCI were used by some schools, but are not described in this paper. In both cases, too few students used the test for meaningful cross-school results to be given. It should be noted that the FCI (only rarely used in this project) is not designed for a cluster analysis.[18,19] The data from the FCI are consistent with those from the FMCE described below.

Findings on conceptual tests in research-based curricula (RBCs) such as RTP/ILD have shown that students perform much better in the RBCs than they do in the traditional physics classroom. This holds true both for curricula that bring wholesale change to the classroom (such as Workshop Physics from Dickinson College) and for curricula that change only parts of the classroom (such as Tutorials in Introductory Physics from the University of Washington, Seattle, or Group Based Problem Solving, from the University of Minnesota, Minneapolis).[20] We expected that student performance will be on par with performance in the small-scale change RBCs.

B. Expectations Survey

For the MPEX, we also used matched pre- and post-instruction data. Responses on the MPEX were evaluated for being favorable or unfavorable, as matched to expert instructors and the attitudes that they believed would be most helpful for students learning physics.[12] Though the MPEX is designed on an agree/disagree, it is sometimes favorable to agree, sometimes favorable to disagree, and so on. Thus, evaluating the test in terms of whether an answer was deemed favorable in helping students learn the physics seems most reasonable. The percent of student responses that are favorable or unfavorable before and after instruction were compared to see the effect of instruction on student expectations.

As with the FMCE, the MPEX can be broken into clusters of responses. These deal with student attitudes toward coherence of the material, conceptual understanding, independence in learning (as opposed to authority driven learning), the role of mathematics in learning physics, the link between the physics and the "real world," and the effort required to succeed in the course. Each of these clusters has specific questions that can be evaluated as favorable or unfavorable. For a given class, it is possible to analyze the overall MPEX score and the scores on each cluster.

Two trends are important when comparing RTP/ILD data on the MPEX survey to data from other schools and forms of instruction.[20] First, student expectations at the beginning of a class are generally less favorable than instructors would like. Pre-instruction scores are typically between 40% and 60% favorable, regardless of the type of instruction or the type of university. Second, scores generally show degradation during a semester. Students have less favorable attitudes toward the course at the end of the instruction than they did at the beginning. These results are true in all lecture/laboratory/recitation format classes, even when parts of the curriculum have been replaced by research-based curriculum materials.[12]

Only in one university instructional setting have increases in scores been observed.[20] In some university settings of Workshop Physics, a well-integrated studio physics curriculum developed by Priscilla Laws, student attitudes improved. Other improvements have been found in instructional situations (e.g. at the high school level[21] and certain experimental university classes) in which student epistemologies were a primary focus of the course.

Since the RTP/ILD materials were implemented in a mostly traditional lecture/lab/recitation setting, student scores were expected to match those in similar settings. Since RTP/ILD involves far more live data gathering and the use of real data in a more easily interpretable form, we hoped that the RTP/ILD materials would lead to improved results in the cluster of responses showing a link between the physics and real world.

III. Testing Results

In reporting data on student learning in the physics classrooms using RTP/ILD materials, we will report general trends in the data, rather than pulling out individual descriptions of each school. In the process, we wish to emphasize the general nature of success (or failure) for the project as a whole. Individual schools with noteworthy results will be discussed in detail as appropriate.

A. Conceptual learning

We describe student learning as measured by conceptual tests administered before and after instruction on a given topic.

Conceptual Test Data: FMCE

Several consistent results have been found in the evaluation of student understanding of force and motion. These are:

- Students in RTP/ILD classes consistently perform better than students in traditional instruction classrooms with gains from 0.30 to 0.70 as compared to gains of 0.1 to 0.35,
- Students in ILD-only classes can perform as well as RTP-only and RTP/ILD classes, but these classes are far more (negatively) sensitive to implementation issues,
- Classes taught in a studio format (but using RTP materials) score higher than any other form of instruction on conceptual tests.
- Implementation issues (e.g. incomplete, unsupported, or with weak instructor support) can have a serious effect on both ILD and RTP classes, such that student learning is no different from traditional instruction classes.

Figure 1 presents the data in summary form. Several classes are noteworthy in this set of data.

ILD-only instruction

One of the high-gain ILD-only courses was taught in a three week winter term at School A (Winter, 1999). We would not expect students to achieve any major gains in such a short course. Conversely, effects of "test memory," meaning improvement in scores due to recall of questions from the previous time the test was taken, may play a role here. Still, it is encouraging to see that the ILD results in such a setting can be at the top end of the traditional spectrum.

In comparison, two of the low-gain ILD-only scores and one of the high-gain ILD-only scores were from School B and illustrate the difficulties in getting good results from ILD only instruction. The low-gain classes had problems with low class attendance, very bad student attitudes toward the materials, and problematic implementation of the materials. The high-gain class, taught in a different semester, had a special focus on attendance and correct use of the materials in the classroom. These data show that implementation issues with ILD materials are much more important than with RTP materials, since the materials are only rarely used, student attendance is central, and the demonstrations may meet with strong objections from the students who are unfamiliar with this mode of instruction during lecture time.

Comparing instructional methods at a single school

To illustrate that comparison across schools while using different curricula is at the very least plausible, see Figure 2. The results in this figure are taken from School C, where three different forms of instruction (traditional lecture and lab, traditional lecture with RTP lab, and studio physics using RTP materials in a lab-only course) were used in any single semester. Figure 2 (a) gives the Winter, 1999, data, while Figure 2(b) shows data from three semesters in a slightly different representation. The students entering the each of the courses scored roughly evenly on the pre-instruction FMCE test (with scores between 26% and 33%). The great difference in their scores comes on the post-instruction results. Here, we find that students in the studio class had normalized gains between 0.52 and 0.65, while students in the traditional-only class had normalized gains between 0.12 and 0.21. The class that used RTP labs in addition to the otherwise traditional lectures scored in between, with a gain of 0.42. These results are roughly consistent with what was found in other studio class settings.[20]

The value of these data for choices in instruction is apparent. This test site is now able to make more informed choices about its instructional methods through a cost/benefit analysis of the two more effective instructional methods. With a studio class, they need more computing equipment, a redesigned laboratory space, and other large capital costs. With the RTP class, they keep the lecture format, while still having conceptual learning gains far above that of the traditional class. As part of the evaluation efforts, such results can have an effect on individual schools, in addition to helping evaluate the overall effectiveness of the materials.

Further analysis possibilities

Two more points can be made with respect to the FMCE data gathered in this evaluation project. They concern different types of analyses that were possible or necessary.

For each school and for each class for which data exist, there is a cluster breakdown for the scores in that class. For example, in Figure 3, the data from School A in the Fall, 1999, semester are given. These show that the class began with relatively low pre-instruction overall score, but had a very good normalized gain of 0.45. When looking at the different clusters, the Newton 3rd Law cluster is noteworthy. An ILD on this topic was carried out in the class in addition to the RTP labs that dealt with the concept. Results from other schools in which ILDs were carried out also indicated that students performed better on these targeted clusters than on others. Since attendance data for each of the ILDs from the different schools were unavailable, we were unable to separate students into two populations which did, or did not, participate in the ILDs, so these results cannot be described in more detail.

A second interesting element of analysis was carried out for data from School D. In these classes, which had strongly varying levels of implementation that were generally undocumented by the school, gains were the same, regardless of the level of instruction (between 0.25 and 0.35 from semester to semester). Due to the lack of information about implementation in the different classes, additional data was sought from School D. Thus, the data from the Fall, 1999, semester were broken down not by instructional method, but according to major. These data, shown in Table 1, indicate that engineers benefited more from the RTP and ILD materials than did science or liberal arts majors. A certain level of self-selection within the student body of the school may be the reason for the differences in normalized gain averages for the majors. Also, it may be that engineering majors at the school are more motivated to succeed in the course, since they are aware that they will need the material later in their studies. School D differs from other schools in this study in that all students are required to take physics. Thus, students who would not have chosen to take physics (and care less about mastering the material) at other schools are included in the data set for School D.

Both these issues raise questions about the success of implementation at test sites. Without additional information, we cannot understand either the specific success of the ILD materials or the affective issues leading to success of certain majors over others when using RTP and ILD materials. Affective issues in general will be discussed below, in the context of the MPEX data.

Conceptual Test Data: ECCE

The data from the Electric Circuits Conceptual Exam are less clear than the FMCE data in terms of a clear distinction between traditional instruction and RTP lab instruction. For an overview of the data, see Figure 4. Note that fewer schools taught electric circuits and gave their students the ECCE, so that the conclusions that can be drawn are based on fewer data points than in the case of student understanding of force and motion as measured by the FMCE. Several additional elements of the evaluation are worth noting.

Highly successful implementation of RTP/ILD materials

Results from School A in classes which used both RTP and ILD materials show that making a clear effort to help students in many ways (both in lab with RTP materials and in lecture with ILDs) can have a strong effect on student learning. Normalized gains in these courses were 0.64 and 0.68. Instructors at the test site reported that they put great effort into making these materials helpful for the students. One should note, though, that the ECCE and the RTP/ILD materials are often very close in design, style, and things such as the figures used, such that students engaged in both RTP and ILD activities are more likely to perform well on the test due to familiarity with representation, for example. The exceptionally high normalized gains indicate that additional factors should be at work, though, since other schools which used both ILD and RTP materials did not have such high conceptual learning gains.

Student learning in a studio physics setting

Three studio physics classes were also evaluated (two at School C, one School E). The studio physics class at School E was taught using Workshop Physics materials which are closely related to the RTP materials, while the School C studio classes used materials modified from the original RTP labs. The data indicate that students in the studio classes perform on average better in the studio classes than in the RTP lab classes.

In the two studio courses at School C, an interesting point can be made. As with their FMCE data, the school was able to have one class of studio instruction and one class of traditional instruction in a single semester. Both of the studio courses had normalized gains of 0.46. One traditional instruction class had a normalized gain of 0.27, the other of 0.40. These data do not show as clear a distinction between studio and traditional as was the case in the FMCE data.

Failure to properly implement materials

Three of the four low-gain RTP classes were from the School D. This school had serious implementation issues, with many classes taught in many different fashions. In general, all used only a subset of the RTP or ILD materials, often combining separate labs into one. The level of implementation in each of these classes was difficult to determine, since some of the materials were rewritten at the test site. In some cases, it is possible that students were only doing a few of the RTP labs as originally designed. Thus, we view this school as an example of the difficulties caused by incomplete implementation of the materials. (Note, also, that this test site had the lowest scoring RTP/ILD classes on the FMCE for similar reasons, though there the results were better than on the ECCE.)

Summary of data

Overall, the ECCE data indicate success in the RTP labs (especially when using ILDs in addition). The data are less compelling than in the case of FMCE data, though. We believe that the ECCE data indicate sensitivity to implementation in the same fashion that the FMCE indicated sensitivity to implementation issues in the ILD-only schools. Among the reasons for this sensitivity are that the RTP labs are taught for more time than is usually given to the topic in instruction. Thus, students are working on the lab materials for longer than they are studying the materials in other parts of their course. As a result, success in implementation plays a serious role in what students learn. Further information and study would be required to understand this issue in more detail.

B. Student attitudes toward learning: Expectations Survey

We expected that students in the RTP/ILD project would show results similar to those found in traditional instruction classes because the general format of the class (lecture/lab/recitation) was not conducive to students rethinking their general approach to learning physics. This prediction is clearly seen in the data, where we see weakly favorable pre-instruction scores and overall degradation in student scores over the course of instruction.

When presenting the data from the MPEX, we have left out the year 1 results. There are several reasons for this. First, the year 1 data used a shortened set of MPEX questions (designed in order to save time in test-taking at the test sites). Comparing year 1 and year 2 data is therefore problematic, since the year 2 data contain more information. Second, the year 1 data are incomplete, with many schools having test-implementation issues. At some schools, the number of students taking both pre- and post-tests (or answering enough questions to be considered a fair representation of their expectations in the course) was so low that the data are not significant. These shortcomings were addressed in the collection of year 2 data, which give a clear representation of student expectations in and attitudes toward the course, allowing discussion of the RTP/ILD dissemination project.

In the data in Figures 5 and 6, two representations are used. Figure 5 gives data from the first semester of instruction (mechanics), while Figure 6 shows data from the second semester of RTP/ILD instruction. In each figure, two charts are given. Chart (a) shows the class averages of pre- and post-instruction student favorable and unfavorable scores. Pre-instruction scores are indicated by open icons, post-instruction scores by filled icons. The most favorable score would be in the upper left corner of the chart (e.g. 100% favorable, 0% unfavorable). The diagonal line in the graph indicates the line of possible scores if no students are neutral on the agree/disagree questions. Thus, a point's distance from this line indicates the average percentage of neutral responses given by the students in a class (neither agreeing nor disagreeing, thus neither favorable nor unfavorable). Chart (b) shows the movement between the pre- and post- instruction scores. All pre-instruction scores are placed at the origin, in effect normalizing all pre-instruction scores to a single value. A vector has been drawn on the figure to indicate the direction of movement in each class. Note that the classes are not labeled by name but rather by type of instruction in each situation. Icons in chart (b) match those in chart (a). In order to save space, only those quadrants which show data points are shown in Figure 5(b).

Two major trends are obvious when discussing the MPEX data. On the one hand, students begin instruction with expectations that are only slightly favorable. Thought the scores are spread out over a large range, students are generally 40 – 60% favorable and 40 – 20% unfavorable as they enter the course. This obviously creates a problem for instructors wishing to help their students develop a deeper understanding of the physics. Weak expectations when entering the course can lead to bad study habits and an incomplete picture of what is meant by physics.

On the other hand, the movement of the classes is, in all but two cases (discussed below), in the fully unfavorable direction. Students end the semester with less favorable and more unfavorable scores than when they entered the class. (Note that these are two separate measures, since it is possible to be less favorable, while being less unfavorable, also; such a change would be indicated by a vector drawn into the third quadrant of the graphs shown in parts (b) of the figures above.) This leads to two different conclusions. First, students are being lead to believe or follow unfavorable study habits, expectations of learning, and attitudes toward the course. This is a negative result, and one that is highly problematic. It is consistent with results from other lecture oriented classes, indicating that the RTP/ILD materials lead to improved conceptual learning (as shown above), but do not lead to improved expectations in the course. Second, students are able to gain a conceptual understanding of the material while having increasingly unfavorable expectations of the material. This indicates that students may not be aware of their learning while they are learning. The question of productive student epistemologies toward learning lies outside the scope of this evaluation report, but is an on-going research question that needs to be addressed in more detail.[13] Specifically, changes to the materials that improve student attitudes toward learning might lead to improved conceptual test scores, suggesting one possible avenue of improvement for ongoing development of the materials.

The two MPEX results that show both an increase in favorable scores and an increase in unfavorable scores (see Figure 5) have anomalies in the test-taking methods and should not necessarily be interpreted as successes. In both cases, the MPEX was given as a take-home test rather than an online test. Previous research[9] has shown that the MPEX is extremely sensitive to the setting in which it is given. In one class, both the pre- and post-test were given in take-home format, but very few students completed both in spite of the incentives which were given by the instructor (such as extra credit, and so on). This indicates that possibly only those personally motivated to take the test and turn it in are included in the data, and this level of motivation may play a large role in the final result. Similarly, in the other class, the post-test was given as a take-home (the pre-test had been given online, but with serious problems in implementation), and possibly only the highly motivated students returned the tests. Also, the problems with the pre-test may have biased many students against giving fair and honest responses to the questions on the pre-test. As a result, we include the data, but give little weight to the results.

C. Summary of Results on Student Learning and Expectations

Based on the data in both the conceptual tests and the expectations survey, it is clear that students are showing improved student learning while not showing an improvement in their attitudes toward the course. Both of these results were expected, and neither is surprising. They indicate that the implementation of RTP and ILD materials was on the whole successful in those areas in which success should have been found, while typical in those areas in which nearly all forms of instruction cause difficulties.

The results are encouraging for several reasons. First, they indicate that students leave the course with a deeper understanding of the material. Thus, the RTP/ILD dissemination project has reached its goal of helping students learn the physics. Second, by combining MPEX data with conceptual test data, we see that students are gaining in their conceptual understanding in spite of their attitudes toward the course. This result confirms results from other studies, cited previously, and suggests that additional student learning could be attained by addressing attitudinal issues. Third, improved student understanding of the physics, as measured by standard tests, is shown to take place at a conceptual level, while the materials being tested involve hands-on laboratory activities in which students might not explicitly be aware of the conceptual learning. Again, the implication is that students are learning without explicit awareness of the skills they are learning, indicating room for possible improvement of the materials.

IV. Implementation Difficulties

Schools that successfully implemented the RTP/ILD materials were similar in their implementation, while each school that had difficulties had unique problems. Of the six schools, three were successful in their implementation (Schools A, C, and E) and will not be discussed in this section. Of the other three (Schools B, D, and F), each will be described in its own fashion, from the viewpoint of the evaluation project. Many of these results have been included in the discussion above, and so will only be summarized at this point.

Incomplete implementation of ILDs in a low-attendance situation

At School B, only ILD materials were used. As has been previously discussed, the ILD-only data on the FMCE indicate that two levels of implementation existed. In two semesters, the scores were very low and the level of learning, as measured by the concept tests and normalized gains, was also very low (the lowest of all measured scores in this evaluation). In the other semester, excellent results were reported. The test site reported that a substantial effort was required to successfully implement the materials in the semester for which high normalized gains were reported. In all semesters, the person carrying out the ILDs was not the instructor in the course.

These data indicate that certain schools will have greater difficulties in successfully carrying out modifications to their instructional format than others. The details of the difficulties at School B are not necessarily a guide to how issues might arise at other schools, but they suggest several possibilities. For example, a school with low attendance at lecture (such as a commuter school, or a non-residential school) will have greater problems with a successful use of ILDs. Also, it may be that doing ILDs with a person who is not the instructor in the course may cause problems due to attendance and motivation in the classroom. The success of implementation in the core class (when high gains were found) contradicts this, though, indicating that the issue is not well understood. This possibility is contradicted by the success of ILDs at the primary implementation sites[8] and may depend, instead, on the manner in which the professor carrying out the ILDs was unfamiliar with the materials.

Non-research-based revisions to RTP materials

As discussed above, the implementation of materials at School D led to lower student gains on the conceptual tests than at other schools. Materials were modified with no care toward research into specific issues in student learning of the material, there were 25 separate instructors implementing these materials inconsistently from section to section, and a lack of communication to the evaluation team about the specific details of each instructor's use of the materials.

As a result of these difficulties, the evaluation team was unable to discern exactly what problems caused a lowering in student understanding of the material. For example, it seems that the use of ILDs at School D had little effect (when comparing classes that used them to those that did not). The data are incomplete because the information provided to the evaluation team by the school was incomplete. (It is even unclear which ILDs were used in general, and not consistently known which were used in which section). In another case, on ECCE data from the spring, 2000, semester, there was no information given about the different levels of implementation of the materials. As a result, there is no way of knowing whether the relatively low ECCE normalized gains were due to implementation issues, which instructors did or did not use the materials, and how the materials were used.

Poor communication with project leaders as problems occur

There were difficulties in implementing the materials at School F, in addition to problems in evaluating student learning. The implementation issues included incomplete use of materials, defective equipment, and a general lack of communication about problems in the course with the dissemination and evaluation team. In addition, data were gathered during voluntary lab periods, which a majority of the class avoided. As a result, some of the data from School F are left out of this paper. Too few matched students existed for a valid comparison of pre- and post-instruction scores to be made. In other semesters, data sets were incomplete and also had to be dropped. Though the data exist for a cross-semester study (in which all matched students are combined into a single set of data), this is statistically not viable due to the inconsistent implementation of the curricula in each situation.

As a result, the evaluation team cannot point to the success or failure of the implementation at the test site. In general, the data do look favorable, but there is too little information for a complete picture to be drawn. We point this out in order to emphasize that there are two different levels of implementation difficulties on a project such as this one. (This was also clear from the School D data.) The first is when the materials themselves are implemented badly. The second is when the evaluation of the materials is not carried out well. Since the evaluation team relied on the test sites to provide data, this was a weakness in the evaluation plan, but one that was overcome at most of the schools.

V. Conclusions

RealTime Physics and Interactive Lecture Demonstration materials were used at six very different schools during the three years of this dissemination project. In all but those specific cases outlined above, student learning showed great improvement. While student expectations and attitudes toward learning physics did not improve, the move toward unfavorable expectations was not out of line with other, similarly taught, lecture-lab-recitation courses. As a result of these two points, we believe that the project met its goals and should be counted as a success.

We emphasize that improved student learning depends on the success of the implementation of the materials as originally created. Those schools that did not use the correct materials, or modified them without insight into student reasoning or understanding of the topics addressed in a specific activity, had lower scores on the standardized tests. Though all teachers have a desire, at some level, to modify materials to meet their needs, they should be discouraged from doing so without guidance from the developers or others familiar with the research-based nature of the curriculum.

Acknowledgements

The work in this paper was funded in part by FIPSE grant P116B980003. I wish to thank the project leader, David Sokoloff, for his guidance in implementing the evaluation at the different schools and helping gather information about implementation issues at the different schools. This work was carried out while I was a member of the Physics Education Research Group at the University of Maryland, and I thank the group members for help in interpreting the data and devising the analysis templates used in the study. Specifically, Apriel Hodari and Jeff Saul were instrumental in developing the MPEX analysis template. Also, Ron Thornton provided help in developing the FMCE analysis template.

FIGURES

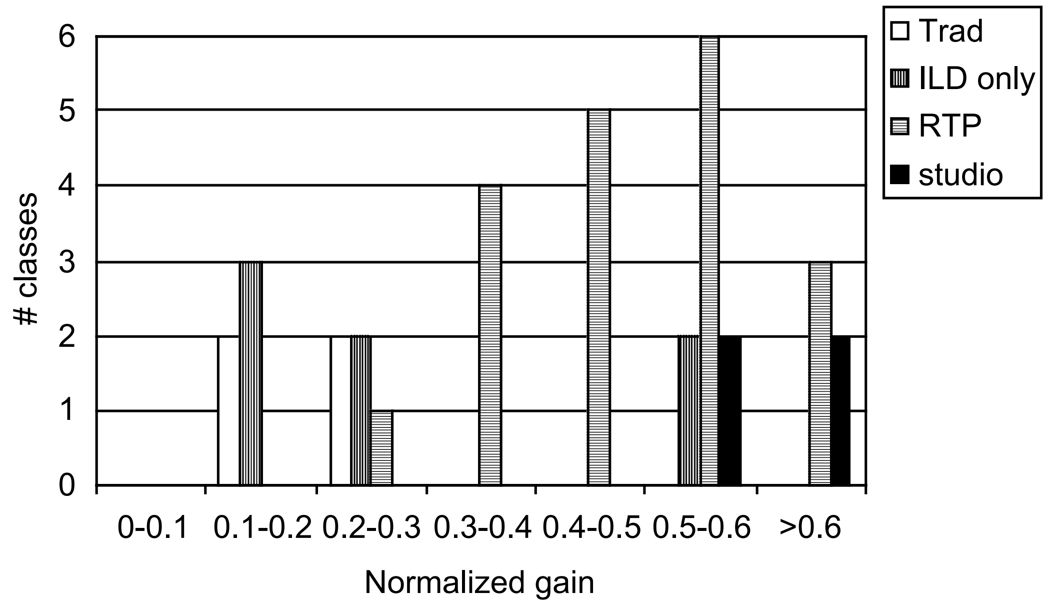
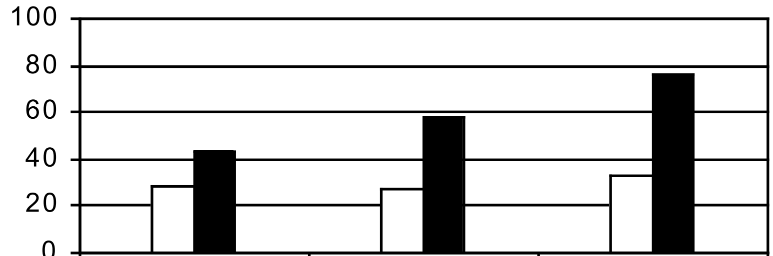


Figure 1: Overview of all FMCE data for all schools. Classes at a single school in a given semester have been combined into a single class, on the condition that the form of instruction in those classes was similar. Higher normalized gains indicate greater improvement on the conceptual test.

(a)



	traditional lecture/lab	Lecture w/ RTP	Studio w/ RTP
N=	87	42	149
□ Pre	28.2	27.7	32.9
■ Post	43.3	57.9	76.7
Gain	0.21	0.42	0.65

(b)

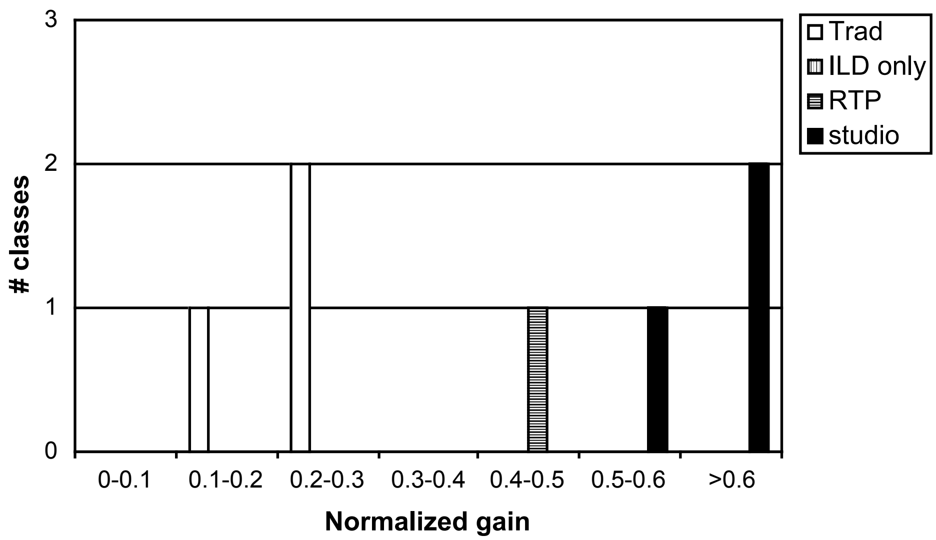
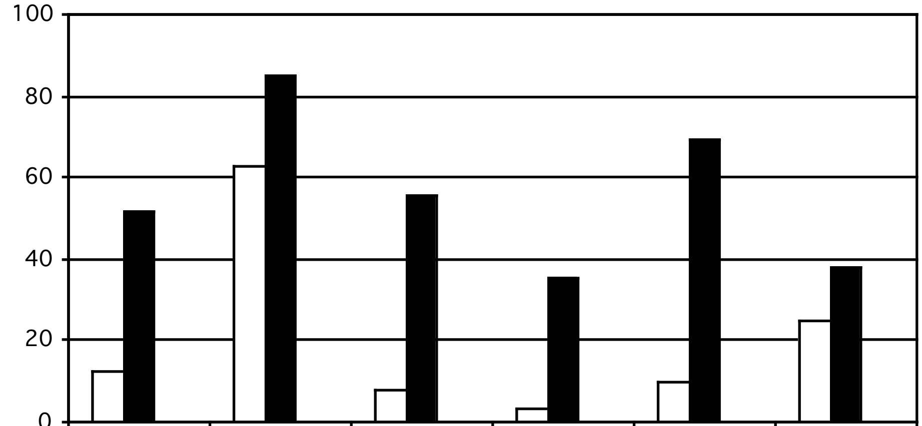


Figure 2: Comparison of FMCE data at one test site. (a) Pre and post scores from the Winter, 1999, semester, with normalized gain data shown. (b) All data from semesters in which both studio and traditional instruction occurred (including that shown in (a)). Normalized gains in were roughly 0.40 higher for the studio than for the traditional instruction class.



	Overall	Velocity	Accel	Forcce (1,2)	Force (3)	Energy
□ Pre	12.6	62.5	7.9	3.3	9.6	25
■ Post	51.6	85	55.7	35	69.6	38.1
Gain	0.45	0.6	0.52	0.33	0.66	0.18

Figure 3: Example cluster breakdown on the FMCE for an individual school (Pacific University, Fall 1999, N = 40 students, using both RTP and ILD materials). Note that a low pre-instruction score on the overall score (using the Thornton rubric) still led to a large normalized gain.

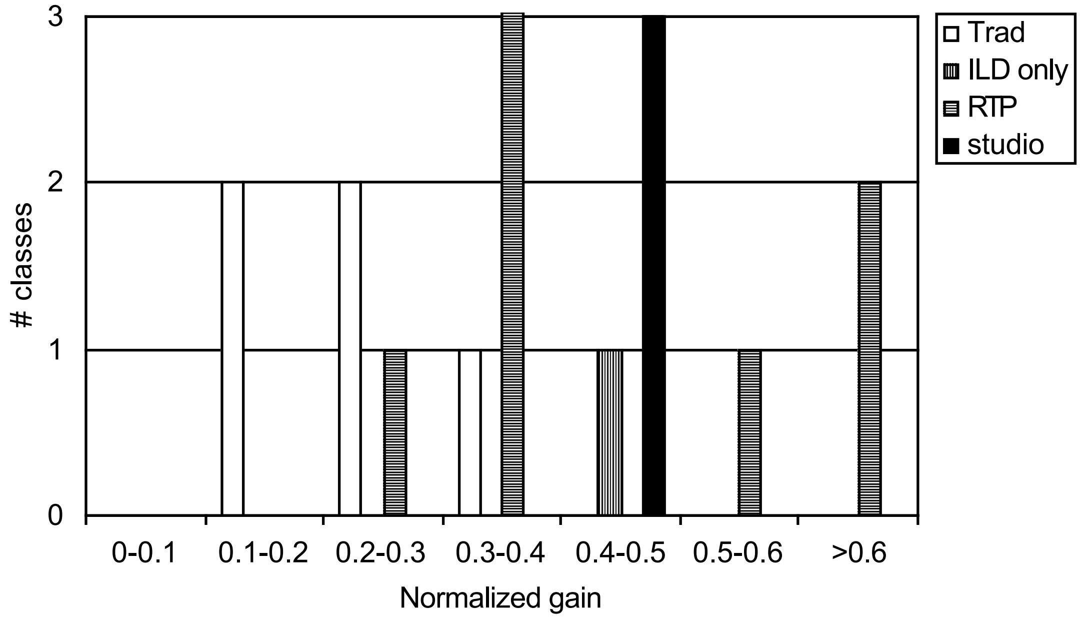


Figure 4: Overview of ECCE data for all schools. Classes at a single school in a given semester have been combined into a single class in this histogram, on the condition that the form of instruction in those classes was similar. Higher normalized gains indicate greater improvement on the conceptual test.

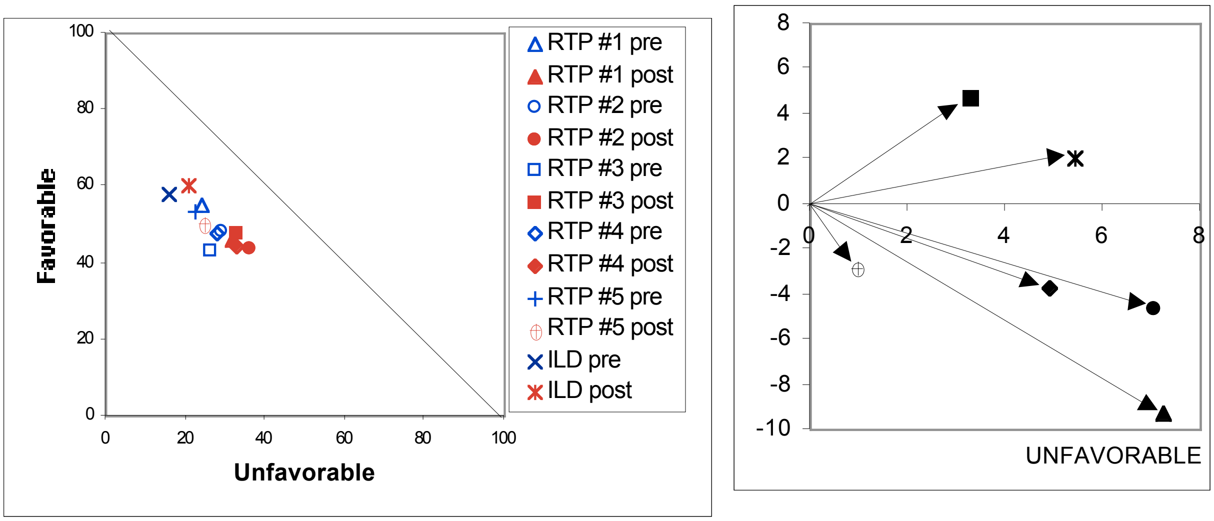


Figure 5: MPEX data from Year 2 of RTP/ILD implementation, mechanics instruction. a) Pre- and post-instruction scores from 6 classes teaching first semester physics, including force and motion. b) The change in scores of classes from their pre-instruction scores. The origin is taken as the pre-instruction score for each school (i.e. normalized across schools).

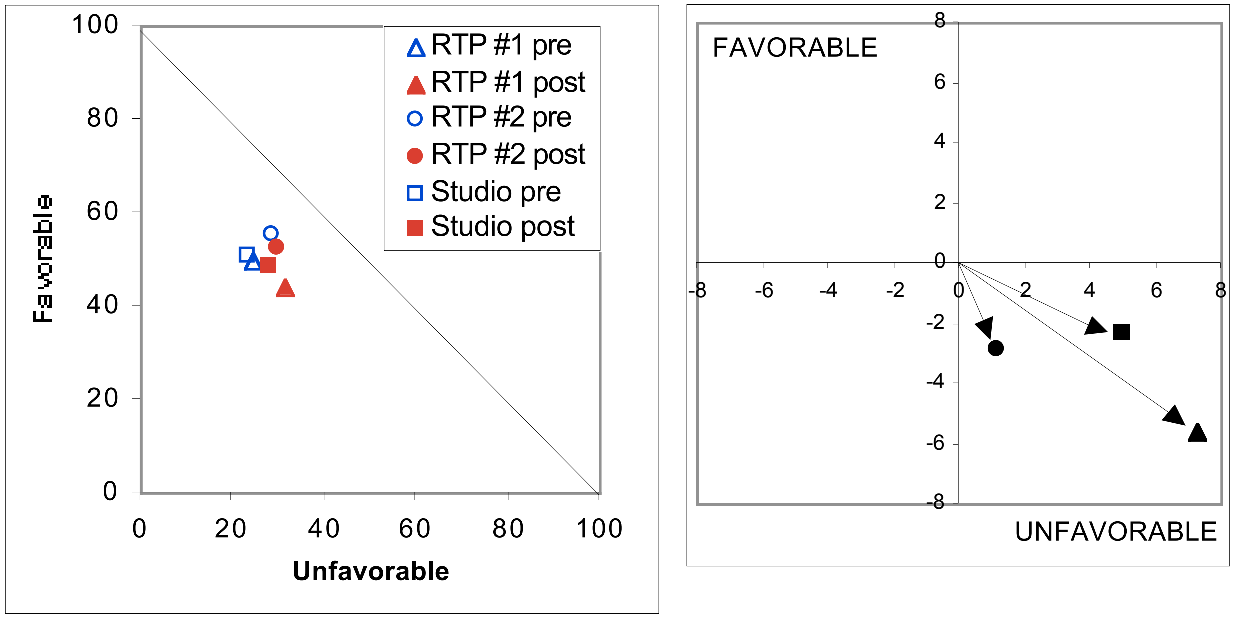


Figure 6: MPEX data from Year 2 of RTP/ILD implementation, electric circuits and heat and temperature) instruction. a) Pre- and post-instruction scores from 3 classes teaching second semester physics, including electric circuits and heat and temperature. b) The change in scores of classes from their pre-instruction scores. The origin is taken as the pre-instruction score for each school (i.e. normalized across schools.

TABLES

Major	Normalized gain
Engineering	0.39
Science	0.31
Liberal Arts	0.29

Table 1: Data from the United States Naval Academy, in which normalized gains on the FMCE for all forms of instruction were very similar (between 0.25 and 0.35), but where Fall, 1999, data indicate that some majors were gaining more from the use of RTP and ILD materials than other majors.

ENDNOTES

[1] Lillian C. McDermott and Edward F. Redish, "Resource Letter PER-1: Physics Education Research," American Journal of Physics 67, 755-767 (1999).

[2] Lillian C. McDermott and Peter S. Shaffer, "Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding," American Journal of Physics 61, 994-1003 (1992).

[3] Lillian C. McDermott and Peter S. Shaffer, "Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of an instructional strategy," American Journal of Physics 61, 1003-1013 (1992).

[4] Lillian C. McDermott, Peter S. Shaffer, and Mark D. Somers, "Research as a guide for teaching introductory mechanics: An illustration in the context of the Atwood's machine," American Journal of Physics 62, 46-55 (1994).

[5] Edward F. Redish, Jeffery M. Saul, and Richard N. Steinberg, "On the effectiveness of active-engagement microcomputer-based laboratories," American Journal of Physics 65, 45-54 (1997).

[6] Richard N. Steinberg, Graham E. Oberem, and Lillian C. McDermott, "Development of a computer-based tutorial on the photoelectric effect," American Journal of Physics 64, 1370-1379 (1996).

[7]Ronald K. Thornton and David R. Sokoloff, "Learning motion concepts using real-time, microcomputer-based laboratory tools," American Journal of Physics 58, 858-867 (1990).

[8]Ronald K. Thornton and David R. Sokoloff, "Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active

Learning Laboratory and Lecture Curricula," American Journal of Physics 66 (4), 338-352 (1998).

[9]Jeffery M. Saul, "Beyond Problem Solving: Evaluating Introductory Curricula Through the Hidden Curriculum," Ph.D. dissertation, University of Maryland, 1998.

[10]Richard R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," American Journal of Physics 66, 64-74 (1998).

[11]I Halloun, "Views about science and physics achievement: The VASS story," AIP Conf. Proc. 399, 605-613 (1997).

[12]Edward F. Redish, Jeffery M. Saul, and Richard N. Steinberg, "Student expectations in introductory physics," American Journal of Physics 66, 212-224 (1998).

[13] David Hammer and Andrew Elby, "On the form of a personal epistemology," in *Personal epistemology: The psychology of beliefs about knowledge and knowing*, edited by B. K. Hofer and P. R. Pintrich (Lawrence Erlbaum Associates, Mahwah, NJ, 2002).

[14]Project Galileo, "Project Galileo," http://galileo.harvard.edu/index.html.

[15]Karen C. Cummings, Jeffrey Marx, Ronald K. Thornton et al., "Evaluating innovation in studio physics," American Journal of Physics 67 (supplement 1 to no. 7), S38-S44 (1999).

[16] Lei Bao, "Mathematical issues in using score gain to assess student performance," http://www.physics.ohio-state.edu/~lbao/archive/papers/g-factor_pre.pdf, (in preparation).

[17]Ronald K. Thornton (Personal Communication).

[18]P. Heller, Huffman, D., "Interpreting the force concept inventory. A reply to Hestenes and Halloun," The Physics Teacher 33 (8), 503,507-511 (1995).

[19]D. Hestenes, Halloun, "Interpreting the force concept inventory. A response to march 1995 critique by Huffman and Heller," The Physics Teacher 33 (8), 502,504-506 (1995).

[20]Jeffery M. Saul and Edward F. Redish, "An Evaluation of the Workshop Physics Dissemination Project, FIPSE Grant #P116B50026,", (1998).

[21] Andrew Elby, "Helping physics students learn about learning," American Journal of Physics 69 (Physics Education Research Supplement) (7 (supplement 1)), S54-S64 (2001).