

Real Time Physics and Interactive Lecture Demonstration Dissemination Project - Evaluation Report

Michael C. Wittmann
Department of Physics
University of Maryland
College Park MD 20742-4111

I. Introduction

The Real Time Physics/Interactive Lecture Demonstrations (RTP/ILD) dissemination project seeks to implement the RTP and ILD curricula at a variety of schools, ranging from community colleges to research universities.¹ The goal is to provide research-based curriculum materials that have been proven to provide effective instruction to students whose physics understanding is shown to be sorely lacking when they leave the traditional classroom. This report describes how the implementation is evaluated and tested at the different test sites.

When investigating the effectiveness of the secondary implementation of the RTP/ILD curricula, the primary focus has been on student learning. For a curriculum to be considered effective and relevant to the greater community, it must first prove to be effective and relevant for the students who occupy our classrooms. Though the RTP/ILD curricula has shown great promise at its primary, developmental site,^{2,3} the open question remains whether the secondary implementation can be as effective and whether student learning shows gains as striking as at the primary site.

This report outlines how we are investigating issues of student learning, student attitudes toward the course, the effectiveness of instructor implementation, and the correlation between student learning, student attitudes, and the level of implementation of the test curriculum. The first part of this report describes the research techniques used to gather data about the implementation of RTP/ILD at the test sites. The second part describes the manner in which the data are analyzed. A later report will complete this preliminary paper by describing the data that we have gathered and what it tells us about the effectiveness of the dissemination of Real Time Physics.

II. Data collection

Five conceptual tests and one attitudinal survey are the primary tools for gathering data about student performance in the implementation sites. In addition, instructor comments are gathered through the use of surveys and two different types of interviews. In this section, each of these data sets is described in more detail.

A. Investigations of Student Performance

To assess the effectiveness of student learning, we use a variety of conceptual tests and a survey that measures student expectations in the course. We then compare the results gathered in pre- and post-testing to a large database of previous studies using a similar format. Exact methods of data analysis will be described in a subsequent section; in this section I describe how data are gathered.

Conceptual Tests

Research has shown that students have deep conceptual difficulties with the material they learn in the physics classroom.⁴ The primary goal of the RTP/ILD classroom is to improve student learning of physics. Improved understanding of physics is greatly needed in the community at large, as most students in the RTP curriculum are taking physics as a service course for other majors, such as biology, engineering, and the pre-med studies. To investigate student understanding, we use a variety of previously developed conceptual tests. Because the conceptual tests have already been validated and tested for reliability to varying degrees, no interviews will be carried out with students at the test sites. Instead, the data on conceptual understanding will come solely from the diagnostic tests.

Each of the five conceptual tests used in this study has a long background and has been used to test a great variety of schools and their instruction in physics. The five tests are the Force Concept Inventory (FCI), Force Motion Conceptual Evaluation (FMCE), Electric Circuits Conceptual Evaluation (ECCE), Heat and Temperature Conceptual Evaluation (HTCE), and Light and Optics Conceptual Evaluation. They are validated tests that have been developed through a long process of research into student difficulties, interviews to validate the questions of the tests, and revisions based on findings during the validation period. Presently, a large database of student results exists for each of the tests.⁵

The conceptual examinations are all included in the appendix of this report.

Expectations Survey

A second type of survey test is 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.^{6,7}

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.

The MPEX survey is included in the appendix to this report.

Gathering Student Data

Two methods of data collection are being used during the evaluation of the RTP/ILD project.

First, students are taking pre- and post-instruction conceptual surveys during classroom time. Instructors use either lecture or laboratory time to give the students the conceptual tests. These are then sent to the University of Maryland, where they are scored and evaluated. Surveys are administered both before and after instruction. Only those students who answered both before and after instruction are compared (we refer to this as matched data). 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.

A second method of data collection has been used with the MPEX survey. Students answer these questions outside of the classroom. The Galileo team at Harvard University has placed the MPEX survey on-line in a form accessible on the World Wide Web. Students can log in on their own time and take the survey as they see fit within a time window specified by the instructor. Scores are compiled at the Harvard Galileo site and then sent to the evaluation site at the University of Maryland.

A slight modification of the outside-class administration of the MPEX is used at two schools in the dissemination project. Students are given the scantron sheet (as is used with in-class testing) and the actual test to take home. These are due in the next class. As with the on-line administration of the test, students are free to take it at any time they like. These tests are then evaluated just like the other in-class forms that are collected from the students.

Both forms of out-of-class testing were introduced due to limitations of the original MPEX testing in the first year of RTP implementation. A short version of the MPEX survey 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. 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 many other variables held constant with respect to the RTP/ILD implementation. Unfortunately, this is not possible, for various logistical reasons. Fortunately, the data have shown that in general, student performance in various classrooms is accurately described regardless of instructor, setting, and a variety of other settings. As has been shown by Hake⁸ and Saul,⁹ interactive engagement classrooms provide a better setting for student learning than traditional instruction classrooms do. But, as Karen Cummings et al.¹⁰ have pointed out, the interactive engagement activities must be implemented correctly, otherwise large expense leads to little result.

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. 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.¹¹

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.

With the three different data sources, it is possible to evaluate the spectrum of possible levels of implementation, from straight traditional to fully implemented RTP curricula.

B. Level of Implementation

A variety of methods of assessing the level of implementation are used in the RTP/ILD dissemination project. These include site visits by the project leader and evaluation leader. Interviews are carried out by the evaluation leader during the semesters in which the team leaders are teaching from the RTP/ILD materials. Further interviews are done during the FIPSE team meeting halfway through the dissemination project. Written reports are to be submitted by the team leaders at the end of the grant period. Finally, survey questions have been developed to help gain insight into implementation issues that may not be visible or noticeable in the other data sets. In each case, we hope that team leaders will describe to us implementation issues, provide us with commentary on the materials, and describe issues that the project and evaluation leader can address in coming semesters.

Site Visits to Test Locations

Each year of the dissemination period, the project leader will visit the schools at which RTP/ILD is being implemented. During these visits, the project and team leaders share views, and the project leader observes at least one classroom in which the dissemination materials are being implemented. These site visits have been instrumental in the design of more appropriate and thorough interaction between the project leader and the test sites. Detailed reports of the site visits are instrumental in determining the level of implementation at each test site. This information is important in analyzing the data describing student performance in the course.

Telephone Interviews with Team Leaders

A second way to gather information about the level of implementation is through phone interviews of the team leaders and instructors during the teaching semester. By getting feedback while the instructors are in the middle of teaching the materials, it is possible to learn about issues that may be forgotten with time. Furthermore, it allows an outside observer to see which issues play a role in instructor acceptance of the material. The possible time development of instructor attitudes also can be measured.

Presentations of and Interviews with Team Leaders

A research team meeting has been organized for October, 1999, in Oregon. During this meeting, the team leaders will make presentations that describe issues in implementation at their schools. These will include both positive and problematic aspects of the implementation.

Surveys on Implementation and Evaluation Issues

During the Oregon meeting, 1999, team leaders will also be given a set of surveys to fill out. These surveys ask team leaders to discuss two different sets of issues. The first is the role of the diagnostic testing in the classroom. We would like to know how the test sites react to the data on student learning and student expectations, and how the test sites react to the loss of class

time. In the second survey, team leaders are asked to describe implementation issues at their schools. This survey serves two purposes, to promote team leader awareness on certain implementation issues, and to gain information that may possibly not be contained in the interviews carried out at other times.

The surveys are included in the appendix to this document.

Team Leader Final Reports

The final source of information about the level of implementation at the test sites will be in the final reports submitted by the team leaders at the test sites. In these reports, they will describe implementation issues, difficulties, successes, and more. These reports will serve a fundamental role in assessing the final attitudes of the instructors toward implementing a research based curriculum not developed at their school.

II. Data Analysis

In this section, the manner in which the many different sorts of data will be interpreted is described. The section begins with a description of the analysis of the conceptual test data, proceeds to the expectations survey data, and ends with a discussion of the role of the instructor data in interpreting the results.

A. Analyzing Student Performance

Previous research results provide methods with which to analyze student conceptual test and expectations survey responses.

Conceptual Test Data

The different conceptual tests all have well accepted methods of analysis that can be applied to them. Presently, student scores from before and after instruction are analyzed at the University of Maryland. Only matched students are included in the sets of data. Once scores have been evaluated, the scores from before and after instruction can be compared.

The comparison is done using the gain factor. "Gain" describes the change in possible score. Thus, a class whose FMCE score went from 20% to 40% correct would have a gain of 0.25 (having moved 20% of a possible 80%, $.20/.80 = 0.25$). Gain factors have become an accepted measure of diagnostic testing because they measure a class's performance against its own standards. Results have shown that schools with similar types of instructional levels have similar gain levels even when the pre- and post-test scores are different.¹² Thus, the gain gives a reasonable and meaningful description of the success of a form of instruction.

Further analysis of the data is also possible. For example, some schools have used both FCI and FMCE testing on the same population. This allows a comparison of the two tests and a further comparison of the school's data to the literature of previous results.

Findings on conceptual tests in research based curricula (RBCs) such as RTP/ILD have found that students perform much better in the RBCs than they do in the traditional physics classroom.¹³ 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).¹⁴

We expect that student performance will be on par with performance in the small-scale change RBCs.

Expectations Survey

The expectations data will be analyzed with pre- and post-instruction comparisons of matched students, also. The responses on the MPEX are analyzed according to whether they are favorable or unfavorable. The percent of student responses that are favorable or unfavorable are then compared to see the effect of instruction on student expectations. Only matched students are included in this analysis, since those who dropped the course early may have had a negative effect on the pre-instruction scores. This does remove a population of students who joined the classes late or did not take the pretest for other reasons, but such a loss of data is unavoidable.

MPEX results show certain trends, much like the conceptual tests show relatively similar types of gains for similar types of instruction.¹⁵ In general, student expectations at the beginning of instruction are less favorable than we would like them to be. Compared to how instructors say they would like students to answer, students only show an overall favorable score between 40% and 60%. This preliminary score is generally independent of instruction, as is to be expected when our classes have not had an effect on student beliefs. It also seems that student self-selection into different level courses (algebra or calculus based, for example) is less visible in their expectations than it is in their incoming conceptual understanding. A second trend that is visible in the data is the general degradation of student expectations over the course of instruction. Student scores become less favorable with time. 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.

Only in one university instructional setting have increases in scores been observed. This was in Workshop Physics, a well-integrated studio physics curriculum developed by Priscilla Laws. Other improvements have been found in instructional situations in which student epistemologies were a primary focus of the course.

Since the RTP/ILD materials are implemented in a mostly traditional lecture/lab/recitations setting, we expect student scores to match to those in other such settings. Since RTP/ILD involves far more live data gathering and the use of real data in a more easily interpretable form, it may be that the RTP/ILD materials will lead to results different from those previously observed in other similar settings.

Instructor Comments

To properly interpret the results of both the conceptual and expectations testing at each implementation site, the level of implementation must be accurately described. For example, the number of students in the various labs must be known, the manner in which pre-labs were administered and homeworks graded must be known, and other issues in student responses should also be known. These data are gathered through surveys, site visits, and interviews with the instructors. By matching their descriptions to the data gathered, we are able to more accurately describe the success of the course. Without clear commentary from the instructors, the other data that are gathered are meaningless and uninterpretable.

III. Conclusion

The focus on the RTP/ILD dissemination project evaluation is on student learning and student attitudes toward the physics classroom. Five conceptual tests and one expectations survey are being used, all of which have been developed on the basis of research into student understanding of the topic. These tests are based on multiple investigations into student understanding of physics. Two have been published in refereed journals. Their development has been documented, and they are in use in the wider physics education community. Results from the tests are reasonably well understood and interpretable. In conjunction with the testing of students, we also ask the instructors for information that helps us describe the course in more detail. These two data sets combined give a detailed understanding of the types of instruction the students had and the effect that this instruction had on their learning and their attitudes toward learning physics. Successful implementation of the RTP/ILD materials would show that student learning is improved in the modified setting, and that trends in student expectations are no worse than in other similar courses. These data results will be reported in a more complete, final report on the dissemination project.

¹ The schools involved are the University of Massachusetts, Dartmouth; California Polytechnic State University, San Luis Obispo; Salt Lake City Community College; the U.S. Naval Academy; Pacific University; and Hunter College of the City University of New York.

² As an example, see R.K. Thornton and D.R. Sokoloff, "RealTime Physics: Active Learning Laboratory," AIP Conf. Proc. 399, 1101-1118 and references cited therein.

³ See R.K. Thornton and D.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," Am. J. Phys. 66, 338-352 (1998) and references cited therein.

⁴ For an overview, see L.C. McDermott and E.F. Redish, "Resource Letter: PER-1: Physics Education Research," Am. J. Phys. 67, 755-768 (1999).

⁵ See reference 3 specifically and 2 additionally for more details on the use of these conceptual tests.

⁶ H. Halloun, "Views about science and physics achievement: The VASS story," AIP Conf. Proc. 399, 605-613 (1997).

⁷ E.F. Redish, J.M. Saul, and R.N. Steinberg, "Student expectations in introductory physics," Am. J. Phys. 66, 212-224 (1998).

⁸ R.R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," Am. J. Phys. 66, 64-74 (1998).

⁹ J.M. Saul, "Beyond Problem Solving: Evaluating Introductory Curricula Through the Hidden Curriculum," dissertation, University of Maryland, College Park, 1998.

¹⁰ K.C. Cummings, J. Marx, R.K. Thornton, and D. Kuhl, "Evaluating innovation in studio physics," Am. J. Physics, 67 (supplement 1 to no.7), S38-S44 (1999).

¹¹ See references 2 and 3 for additional reported scores. Also, personal communication of unpublished scores of other institutions and classrooms come from D.R. Sokoloff and R.K. Thornton.

¹² See, for example, reference 8 for more details.

¹³ For more details, see reference 4 for an overview of many different RBC's and their effects on student learning.

¹⁴ See J.M. Saul and E.F. Redish, "An Evaluation of the Workshop Physics Dissemination Project," FIPSE Grant #P116P50026 (1998) for more details.

¹⁵ See references 7 and 9 for more details.