Undergraduate Physics Course Innovations and Their Impact on Student Learning

Heidi L. Iverson*, Derek C. Briggs*, Maria A. Ruiz-Primo[†], Robert M. Talbot III*, and Lorrie A. Shepard*

*School of Education, University of Colorado, Boulder, CO 80309-0249, USA

[†]School of Education & Human Development, University of Colorado, Denver, CO 80217-3364, USA

Abstract. This paper presents results of an NSF project in which the goal is to provide a synthesis of research on instructional innovations that have been implemented in undergraduate courses in physics. The research questions guiding the project are: What constitutes the range of principal course innovations that are being implemented in undergraduate physics courses? What are the effects of these course innovations on student learning? The paper describes: (1) the literature search procedures used to gather over 400 innovation-related journal articles, (2) the procedures followed to analyze the studies within these articles, (3) the characteristics of the studies reported, and (4) the results from synthesizing the quantitative results of those studies that met our criteria for inclusion.

Keywords: instructional innovations, synthesis, meta-analysis PACS: 01.40.-d 01.40.Di 01.40.Fk 01.40.G- 01.40.gb 01.40.Ha

INTRODUCTION

Over the last several decades, empirical and theoretical research has challenged the efficacy of traditional models of undergraduate physics education, questioning the effectiveness of a single professor lecturing to a group of tens to hundreds of students. Working to improve upon the education of undergraduate physics students the physics education research (PER) community has taken steps to better engage students and help them take a more active role in their own learning. As a result of this growing interest, the number of studies evaluating the effectiveness of instructional innovations [1] has increased considerably; especially after the significant increase in NSF funding in 1991 for these types of projects [2].

The large body of literature on instructional innovations in PER naturally give rise to questions about the effect these various innovations have on student learning. Although there have been several published summaries of PER innovations, these approaches have either been primarily qualitative in nature (making no attempt to synthesize effect estimates across studies) [3] or have been too broad [4] or too narrow [5] in scope with respect to the way that innovations have been operationalized.

This paper describes some of the results of a systematic synthesis of studies that have evaluated the effect of undergraduate physics course innovations. The study is framed by the following research questions:

1)What is the range of PER course innovations?

2) What can be concluded about the effects of these different course innovations on student learning?

A NOTE ON INSTRUCTIONAL INNOVATIONS

Before discussing more of the foundations of this research it is necessary to provide a definition of what is meant by the terms *instructional* or *course innovations*. We use the term undergraduate *course innovations* for those instructional strategies which emphasize an active student learning approach that involves moving away from lecturing as the main or central instructional strategy, shifting the focus from the professor (instructor) towards the student, and supporting shared collaboration [6]. The term *innovations* is used because it encompasses diverse

types of instructional strategies, and it is not associated with any specific strategy. Furthermore, it is the term commonly used within the PER community. The major premise underlying course innovations is that they have been developed with the intention to better facilitate student learning.

PREVIOUS LITERATURE REVIEWS

Several researchers have approached summarizing or synthesizing the body of PER empirical research studies by qualitatively categorizing sub-groups of innovations. For example, Dancy and Henderson [3] developed a framework for evaluating the degree to which classroom activities were more or less "interactive" for students. They evaluated several common PER innovations and provided a narrative synthesis of how these common PER innovations compared to each other. Without the use of comparative data however, the relative effects of these innovations on student learning cannot be established.

Other syntheses summarizing the research literature have been quantitative in nature. Richard Hake's report [4] on the effect of "interactive-engagement" versus traditional teaching methods provided compelling evidence in favor of interactive engagement courses (IE). However, the normalized gains of these IE courses appear to be highly variable, spanning a range of about 0.20 to 0.68. Without more information on the differentiating characteristics of these courses, it is difficult to offer any hypothesis for why some appear to be more effective than others.

Other researchers have approached the task of a PER synthesis in a more fine-grained manner. In this approach, data from multiple replications of a very specific innovation are pooled and averaged. This type of synthesis typically involves replicated studies over the course of several semesters to several years. An archetypal example of this type of synthesis is the work that has been done on analyzing implementations of Peer Instruction [5]. Peer Instruction, developed by Eric Mazur at Harvard University, has been implemented and studied for over ten years. Pre- and post-test student data has been collected over the course of these years and the results show that the average normalized gain for Peer Instruction was 0.61 for calculus-based courses, and 0.64 for algebra-based courses [5]. This is on the upper end of the normalized gain distribution presented in Hake's analysis [4] data. Although this information is informative, very few innovations have been studied extensively in this manner. This does not allow for comparisons to be made across innovations.

In summary, previous reviews of the PER literature have lacked quantitative outcome measures or have either been too broad or too narrow in the way that they have characterized innovations. This study attempts to fill this gap by conducting a quantitative research synthesis that allows for effect comparisons across different instructional innovations. The first research question, *What is the range of PER course innovations?* is addressed with a descriptive synthesis, and the second, *What can be concluded about the effects of these different course innovations on student learning?* is addressed through meta-analysis.

METHODS

The following sub-sections describe the study inclusion criteria and retrieval methods, and the coding framework used.

Inclusion Criteria and Retrieval Methods

Four criteria were established for inclusion of studies in this synthesis study. The study must:

- 1) focus on undergraduate education in physics;
- 2)include one or more instructional strategies considered to be an innovation.
- 3)refer to actual classrooms, rather than controlled conditions; and
- 4) be reported in a paper, article, or document developed or published in 1990 or later [2].

A comprehensive search of research reports on physics undergraduate course innovations was conducted through the process of searching for seminal papers and authors, searching through key journals, and receiving feedback from the project's physics advisory board member and other key researchers in the field. These papers were analyzed to identify those that met the necessary criteria to be included in our study.

From the 414 papers that were collected, 128 were excluded because they did not meet our selection criteria and 286 were kept in the database. Of the remaining 286 articles, 120 were coded as background papers (papers with theoretical information or historical documentation on innovations), 17 were coded as synthesis papers (papers that summarized several studies on one or more innovations either through narrative review or meta-analysis), 32 were coded as descriptive papers (papers that implement innovations but in a non-comparative way), and 118 were coded as research papers with comparative studies involving an experimental design.

Several of the 118 papers reported the results of more than one study within the same paper. Therefore, distinct studies, and not papers, were considered the unit of analysis. Each of these studies represented instances where different treatments, experimental designs, outcome measures, or control groups were used. The final pool included 150 unique comparative studies in undergraduate physics education. Unfortunately, for these comparative studies, only 96 provided sufficient information to calculate an effect size statistic.

Coding Framework

An organizational coding framework was developed by the authors with the intention of categorizing innovations based on their characteristics and the details of their implementations. This organizational framework has been informed and refined through continued use as well as through assistance from this project's advisory board.

The codes primarily described in this paper are those related to distinguishing classes of innovations: (1) *conceptually oriented tasks* represent innovations designed to elicit students' level of understanding of key science concepts, (2) *collaborative learning* represents innovations designed to engage students with groups or in pairs as a component of the learning process), (3) *technology* represents innovations designed to help students visualize processes and/or visualize concepts and/or (4) *inquiry-based projects* represent innovations designed to provide students with the opportunity to undertake research projects for long period of time.

Statistics

As is typical in meta-analysis, the effect size metric is used to place an estimated treatment effect onto a common scale. For experimental designs involving pre- and post-test for an experimental and control group (the design we encountered most frequently), the effect size is computed as

(1)
$$ES = \frac{(\overline{X}_{T_POST} - \overline{X}_{T_PRE}) - (\overline{X}_{C_POST} - \overline{X}_{C_PRE})}{SD_{PRE}}$$

where \overline{X} represents a test score mean for treatment and control conditions (subscripts "T" and "C") administered at the beginning and end of a study period (subscripts "PRE" and "POST"), and SD_{PRE} is computed as the weighted average of the pre-test standard deviations across treatment and control groups.

FINDINGS

Although this research is ongoing, the sections below will describe the findings of this research to date. These findings focus on the results of the descriptive synthesis and meta-analysis.

Descriptive Synthesis

The large majority of studies coded involve *conceptually oriented tasks* (as defined above), either as the primary innovation alone, or in combination with other innovations (e.g., technology). Innovations involving the use of conceptually oriented tasks represent over 65% of the total number of studies in this analysis (N = 150).

The most frequent combination of innovations in this pool of literature involved conceptually oriented tasks combined with collaborative learning and technology. An example of this combination of innovations is Peer Instruction where students use clickers to respond to conceptually oriented questions before and after discussing the questions with their peers [see example in Ref. 7].

The second most common combination of innovations involved conceptually oriented tasks combined with collaborative learning. This combination of innovations is exemplified by the use of Tutorials in Introductory Physics [8]. With Tutorials students work in small groups collaboratively on conceptual questions which have been designed to elicit and resolve students' common alternative conceptions.

The second most common primary innovation is technology which represents 22% of all studies. Technology is most commonly implemented as a singular innovation. An example of this type of innovation is the use of web-based homework like the commonly used CAPA online homework submission program, part of the LON-CAPA course management system (http://lon-capa.org). Here the focus is not so much on the nature of the questions that are asked in the homework, but on how the technology itself can facilitate student learning: for example, through the use of feedback.

Meta-Analysis

Studies to be meta-analyzed are most appropriately combined when they are based upon common experimental designs. It is important to note that virtually all of the studies we reviewed were quasiexperimental in nature: students are not assigned to treatment or control conditions at random. However quasi-experimental designs which involve the use of a pre-test help to reduce (though not eliminate) the inherent threat of selection bias. Out of the overall 96 studies in which we were able to calculate effect sizes, 38 studies fit this criterion, and we summarize the results of effect sizes estimated for these studies in what follows. Figure 1 depicts the average effect size for these studies as a function of possible combinations of innovations under evaluation. The results indicate that the most effective instructional innovation appear to be those that emphasize collaborative learning. The second highest average effect sizes is found for studies in which conceptually oriented tasks are the principal focus in combination with collaborative learning and technology.



FIGURE 1. Distribution of average effect sizes for innovation types: conceptually oriented tasks (COT), collaborative learning (CL), technology (Tech), and inquiry-based tasks (IBP). The numbers shown in white indicate the number of studies in each category, and total size of the underlying student sample is shown in parentheses.

The brevity of this paper limits the degree to which we may further describe the manner in which studies have been further pooled and compared. This information will be included in future manuscripts.

CONCLUSIONS

The results of this research indicate that the effects of innovations in undergraduate physics education on student learning are sizable, albeit variable depending on the type of innovation. However it is difficult to conclude that these results are generalizable. There are three reasons for this:

- 1) The number of studies within groups of similar innovations is small. These numbers will be reduced further when studies are partitioned again into not only similar innovations but similar control groups, and outcome measures, to name just a few factors.
- 2)As noted above, the large majority of effect size estimates within this database derive from quasiexperimental designs (94%) in which issues such as selection bias and attrition are rarely addressed explicitly. It is unclear the degree to which these factors may be compromising the internal validity of the effects estimated in these studies.
- 3) The number of studies reporting effects as a result of systematically developed and tested outcome measures is relatively few (36%). More often, the

effects of innovations are reported based on internally developed outcome measures, such as final exams and quizzes, and there is no reporting of the internal consistency or construct validity of the measure.

These results highlight several methodological characteristics which should be considered in designing future studies: (1) the use of comparison groups, preferably contemporaneous, (2) the use of technically sound instruments, (3) explicitly addressing threats to internal validity such as selection and attrition, and (4) reporting the necessary information to better quantify effects – sample sizes, mean scores, and standard deviation of these scores by group,

ACKNOWLEGEMENTS

We would like to thank Joe Redish and Jose Mestre for their assistance and feedback. This research is funded by NSF grant REC-REESE 0635491.

REFERENCES

- Heron, P.R.L., & Meltzer, D.E. (2005). The future of physics education research: Intellectual challenges and practical concerns. *American Journal of Physics*, 73(5), 390-394; McDermott, L.C., & Redish, E.F. (1999). Resource letter on Physics Education Research. *American Journal of Physics*, 67(9), 755-773.
- 2. Suter L., & Narayanan, H. (2006). A View of Physics Education Portfolio and Research from The National Science Foundation. Paper presented at the Annual Meeting of the American Educational Research Association, April, 2006, San Francisco, CA.
- Dancy, M., & Henderson, C. (2007). Framework for articulating instructional practices and conceptions. *Physics Review Special Topics Physics Education Research*, 3(010103), 1-15.
- Hake, R.R. (1998). Interactive-Engagement versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses. *American Journal of Physics*, 66(1), 64-74.
- see Crouch, C.H., & Mazur, E. (2001). Peer Instruction: Ten Years of Experience and Results. *American Journal of Physics*, 69(9), 970-977; Fagen, A.P., Crouch, C.H., & Mazur, E. (2002). Peer Instruction: Results from a Range of Classrooms. *The Physics Teacher*, 40, 206-209.
- Ruiz-Primo, M. A., Briggs, D., Shepard, L., Iverson, H., & Huchton, M. (2008). Evaluating the impact of instructional innovations in engineering education. In M. Duque (Ed.). Engineering education for the XXI Century: Foundations, strategies and cases (pp. 241-274). Bogotá, Colombia: ACOFI Publications.
- 7. Mazur, E. (1999). Peer Instruction: A User's Manual. *American Journal of Physics*, 67(4), 359-360.
- McDermott, L.C. & Shaffer, P.S., & Somers, M.D. (1994). Research as a guide for teaching introductory mechanics: An illustration in the context of the Atwood's machine. *American Journal of Physics*, 62, 46.