

Impact of Undergraduate Science Course Innovations on Learning

Maria Araceli Ruiz-Primo*, Derek Briggs, Heidi Iverson, Robert Talbot, Lorrie A. Shepard

Despite revealing some positive impacts, studies too often suffer from weak design and inadequate reporting.

At many colleges and universities, the traditional model of science instruction—a professor lecturing a large group of students—is being transformed into one in which students play a more active role in learning. This has been attributed to mounting evidence that traditional lectures, recitations, and laboratory sessions do not guarantee that students develop deep understanding of critical concepts (1–5).

Since 1991, the National Science Foundation (NSF) increased its support to research projects that reconceptualize undergraduate science instruction. The resulting increase of “student-centered” instructional innovations studies raises questions, for example: How do these undergraduate course innovations vary? Do student-centered innovations in undergraduate science courses have a positive impact on student learning? These types of questions motivated a 3-year, NSF-funded research study that intended to characterize undergraduate course innovations described in published journal articles and to quantify their impact on student learning in biology, chemistry, engineering, and physics courses.

We expand upon prior studies (6–8) to consider more types of innovations from more disciplines. Unlike other syntheses, we did not rely on terms used by authors to refer to their innovations, as names may reflect a general term (e.g., technology) with different meanings. Instead, we classified studies based on four, non-mutually exclusive innovations: conceptually oriented tasks (COTs), collaborative learning (CL), technology (TECH), and inquiry-based projects (IBPs) (see the first table). Our categories allowed characterization of innovations in more detail than in previous syntheses (8).

Innovations: Concepts and Methods

Five inclusion criteria were used to screen articles during the literature search: (i) focus on undergraduate education in biology, chemistry, engineering, or physics; (ii)

inclusion of one or more student-centered innovations; (iii) set in a “real-world” regular classroom and/or laboratory environment, as opposed to conducted in an education laboratory; (iv) published in a peer-reviewed journal between 1990 and 2007; and (v) results communicated in English. [See supporting online material (SOM) for details about study methods.]

Of the 868 articles on course innovations gathered, 82, 18, 23, and 74 described at least one comparative study in biology, chemistry,

engineering, or physics, respectively. Comparative studies involve a contrast between students who have and have not received a given instructional innovation (i.e., treatment versus control), making it possible to evaluate the effect of course innovations on student learning. The unit of analysis was “study,” a unique set of data collected under a single research plan from a sample of respondents (9, 10). An article that reports results of multiple innovations, comparison groups, or outcome measures could have multiple studies.

The final pool included 98, 26, 38, and 148 studies in biology, chemistry, engineering, and physics, respectively.

Studies were coded on two dimensions: conceptual characteristics of the innovations, and methodological characteristics of the study designs. This approach permitted us to cluster conceptually and/or methodologically similar studies (10). Periodic agreement checks revealed acceptable intercoder agreement: 89% on average.

Analysis of conceptual characteristics indicated that most studies (69%) involved more than one innovation. The most frequent combination included COTs and CL (26%). Examples of specific, well-known programs that were given this combination of codes included *Tutorials in Introductory Physics* (11) and *Powerful Ideas in Physical Science* (12, 13), in which students work in small groups collaboratively on conceptual questions designed to elicit and resolve students’ common misconceptions. The combination of COT, CL, and TECH codes was also quite common (27%) for established programs of instruction in physics. For example, in *Peer Instruction* (3), students use electronic clickers to respond to conceptu-

FOUR TYPES OF INNOVATIONS STUDIED

Conceptually Oriented Tasks (COTs):

- elicit students’ level of understanding of key science concepts;
- identify students’ misconceptions, “common sense” knowledge at odds with actual concepts that are known to affect students’ learning;
- engage students in conceptual schemes within a topic rather than isolated facts; help students focus on methods of problem representation and approaches to think about the problem components (solving strategies);
- engage students with real-world problems in creative ways that reflect a conceptually integrated understanding of the content

Collaborative Learning (CL) Activities:

- engage students with peers (groups as small as pairs) as a component of the learning process;
- provide students the opportunity to engage in explanations and discussions as they describe their reasoning, interpretations, and solutions to problems.

Technology (TECH):

- helps students visualize processes and/or concepts;
- helps students manipulate variables by collecting and/or analyzing data that can help them understand a concept or a process or solve real-world problems;
- helps students test theories and models with simulated data;
- provides feedback (online homework, such as problem sets).

Inquiry-Based Projects (IBPs):

- provide students the opportunity to undertake research projects (may or may not be in a real-world setting) that require more than one class period to complete;
- require students to develop a procedure and/or plan to complete the project;
- require students to follow a structured procedure and/or plan provided in advance as a framework to complete the project.

Four types of innovation studied.

School of Education and Human Development, University of Colorado Denver, Denver, CO 80217, USA.

*Author for correspondence. E-mail: maria.ruiz-primo@ucdenver.edu

ally oriented questions before and after discussing the questions with their peers. *Peer Instruction* is also used in the discipline of biology with (14) and without (15, 16) the use of electronic clickers. In engineering, the most common innovations involve technology alone (32%), e.g., computer-based simulations from which “virtual experiments” can be conducted as they typically would be in a laboratory (17).

Effect Sizes: Positive, but Variable

An effect size statistic was used to compare instructional innovations on student learning, expressing different outcome measures on a common scale, i.e., the mean effect as a proportion of the standard deviation of the outcome variable (18) (see the second table). Many of the studies did not report summary statistics needed to compute an effect size (means and standard deviations); thus, 46% of 310 studies had to be excluded from the synthesis (the highest percentage in physics, 52%). Only 18 (11%) of the remaining 166 studies involved random assignment of students to treatment or control conditions (experimental design). The most frequent design (89%) can be classified as a “quasi-experiment” (19), in that students were compared across treatment and control conditions, but were not assigned to these conditions at random. Further, 64% of quasi-experimental studies did not include a pretest to establish baseline conditions before the intervention.

The variability of effect sizes within and across disciplines was substantial. The average effect sizes (20) found in biology (0.54) and physics (0.59) were considerably larger than those found in chemistry (0.27) and engineering (0.08). Quality of research design played a clear role, with the mean effect size for randomized experiments (0.26) considerably smaller than for quasi-experiments (0.50).

Caveats and Recommendations

This evidence suggests that undergraduate course innovations in biology, chemistry, engineering, and physics have positive effects on student learning. However, some caveats are in order.

First, as mentioned, almost half of the comparative studies collected for review had to be excluded because they lacked the simple descriptive statistics needed to compute an effect-size estimate. It is unknown whether these results generalize to what would have been found from the excluded studies. Second, it is difficult to rule out plau-

sible threats to the internal validity of most studies, because there are few examples in which students were randomly assigned to treatment and control conditions. To make matters worse, a substantial number of studies fail to administer pretests, making it impossible to rule out preexisting differences in achievement between groups (selection bias) that could artificially inflate or obscure the effect of the innovations. Effect sizes for comparative studies with random assignment are lower than those without random assignment, which indicates that the latter designs are likely to produce inflated estimates of effectiveness. Another important methodological threat to the validity of these studies is a lack of attention to technical characteristics of the instruments used to measure learning outcomes. For example, of the 71 physics studies included, the vast majority (92%) do not provide information about the validity and reliability of the instruments used. Only 3% of studies overall pay attention to these properties that enable a reader to conclude that a given test truly measures what it is claimed to measure.

In the spirit of improving scientific research of instructional innovations, we make the following recommendations, which we view as a joint responsibility of funding agencies, researchers, journal editors, and reviewers.

First, all studies need to include descriptive statistics (sample sizes, means, standard deviations) for all treatment and control groups on all testing occasions. Second, whenever possible, researchers should attempt to randomly assign students to treatment and control conditions. When this is not possible, efforts should be made to demonstrate that the groups are comparable

before the treatment with respect to variables (e.g., prior academic achievement). Finally, researchers should be attentive to the quality of their outcome measures; if measures are not valid and reliable, subsequent interpretations can become equivocal.

Although the poor quality of some research in this field, and the specific shortcomings that commonly undermine studies, have been discussed before, journals continue publishing these types of papers. We are hopeful that our new analyses provide more simple and straightforward emphasis on these critical issues. Experts in experimental research and methodology in education and experts in educational assessment can contribute a great deal to improve research on instructional innovations in science.

References and Notes

1. I. Halloun, D. Hestenes, *Am. J. Phys.* **53**, 1043 (1985).
2. D. Hestenes, *Am. J. Phys.* **66**, 465 (1998).
3. E. Mazur, *Peer Instruction* (Series in Educational Innovation, Prentice Hall, Upper Saddle River, NJ, 1997).
4. L. C. McDermott, E. F. Redish, *Am. J. Phys.* **67**, 755 (1999).
5. L. C. McDermott, P. S. Shaffer, M. D. Somers, *Am. J. Phys.* **62**, 46 (1994).
6. R. J. Beichner *et al.*, in *Research-Based Reform of University Physics*, edited by E. F. Redish and P. J. Cooney [Reviews in Physics Education Research (PER), American Association of Physics Teachers, College Park, MD, 2007]; www.per-central.org/document/ServeFile.cfm?ID=4517.
7. L. Springer, M. E. Stanne, S. S. Donovan, *Rev. Educ. Res.* **69**, 21 (1999).
8. R. R. Hake, *Am. J. Phys.* **66**, 64 (1998).
9. G. V. Glass, *Educ. Res.* **5**, 3 (1976).
10. M. W. Lipsey, D. B. Wilson, *Practical Meta-Analysis* (SAGE Publications, Thousand Oaks, CA, 2001).
11. L. C. McDermott, P. S. Shaffer, *Tutorials in Introductory Physics* (Prentice Hall, Upper Saddle River, NJ, 1998).
12. D. P. Casavant *et al.*, *Powerful Ideas in Physical Science* (American Association of Physics Teachers, College Park, MD, ed. 3, 2001); www.aapt.org/Publications/pips_samples/cover.pdf.
13. L. Liang, D. Gabel, *Int. J. Sci. Educ.* **27**, 1143 (2005).
14. S. Freeman, *et al.*, *Educ. Life Sci. Educ.* **6**, 132 (2007).
15. M. J. Giuliodori, H. L. Lujan, S. E. DiCarlo, *Adv. Physiol. Educ.* **30**, 168 (2006).
16. S. P. Rao, S. E. DiCarlo, *Adv. Physiol. Educ.* **24**, 51 (2000).
17. T. F. Wiesner, Q. Lan, *J. Eng. Educ.* **93**, 195 (2004).
18. Given two studies with the same effect, an effect-size statistic will be larger if one study has more homogenous students (smaller standard deviation) than the other.
19. W. R. Shadish, T. D. Cook, D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Houghton Mifflin, Boston, MA, 2002).
20. We report the mean effect size without weighting by sample size of each study, despite the inferential argument in favor of weighting the effect sizes of studies as a function of their “precision,” which will be highest for those studies involving the largest samples of students. However, such an approach rests on the assumption that each study represents a “sample” from some larger study “population.” We do not feel such an assumption is warranted here, so we avoid the computation of standard errors for inferential purposes. Instead, we prefer to view our effect sizes as descriptive statistics.

Supporting Online Material

www.sciencemag.org/cgi/content/full/331/6022/1269/DC1

EFFECT SIZES OF INNOVATIONS AND COMBINATIONS OF INNOVATIONS				
Innovations	n	Mean	Median	SD
COTs	9	0.47	0.22	0.70
CL	5	0.68	0.69	0.16
TECH	30	0.37	0.38	0.38
COTs+CL	41	0.54	0.54	0.43
COTs+TECH	18	0.41	0.44	0.61
TECH+IBP	9	-0.11	-0.25	0.67
COTs+CL+TECH	39	0.46	0.50	0.50
Other	15	0.84	0.74	0.65
Total	166	0.47	0.47	0.54

“Other” refers to innovation combinations wherein fewer than five studies were combined. See SOM for details.

10.1126/science.1198976