Improved Learning in a Large-Enrollment Physics Class

Louis Deslauriers *et al.*
Science **332**, 862 (2011);
DOI: 10.1126/science.1201783

Acknowledgments. This work was supported by the Belgian Fonds National de la Recherche Scientifique (FNRS), European Commission, Mind Science Foundation, McDonnell Foundation, French-Speaking Community Concerted Research Action (ARC 06/11-340), Fondation Léon Frédéricq, and National Institutes of Health. M.-A.B. and O.G. are Research Fellows, M.B. and C.S. Postdoctoral Fellows, and S.L. Senior Research Associate at the FNRS. M.I.G., V.L., and K.F. are supported by the Wellcome Trust.

Supporting Online Material
www.sciencemag.org/cgi/content/full/332/6031/858/DC1
Materials and Methods
SOM Text
Fig. S1
Table S1
References

22 December 2010; accepted 6 April 2011
10.1126/science.1202043

Improved Learning in a Large-Enrollment Physics Class

Louis Deslauriers,[1,2] Ellen Schelew,[2] Carl Wieman*†‡

We compared the amounts of learning achieved using two different instructional approaches under controlled conditions. We measured the learning of a specific set of topics and objectives when taught by 3 hours of traditional lecture given by an experienced highly rated instructor and 3 hours of instruction given by a trained but inexperienced instructor using instruction based on research in cognitive psychology and physics education. The comparison was made between two large sections ($N = 267$ and $N = 271$) of an introductory undergraduate physics course. We found increased student attendance, higher engagement, and more than twice the learning in the section taught using research-based instruction.

The traditional lecture approach remains the prevailing method for teaching science at the postsecondary level, although there are a growing number of studies indicating that other instructional approaches are more effective (*1–8*). A typical study in the domain of physics demonstrates how student learning is improved from one year to the next when an instructor changes his or her approach, as measured by standard concept-based tests such as the Force Concept Inventory (*9*) or the instructor's own exams. In our studies of two full sessions of an advanced quantum mechanics class taught either by traditional or by interactive learning style, students in the interactive section showed improved learning, but both sections, interactive and traditional, showed similar retention of learning 6 to 18 months later (*10*). Here, we compare learning produced by two contrasting instructional methods in a large-enrollment science course. The control group was lectured by a motivated faculty member with high student evaluations and many years of experience teaching this course. The experimental group was taught by a postdoctoral fellow using instruction based on research on learning. The same selected learning objectives were covered by both instructors in a 1-week period.

The instructional design for the experimental section was based on the concept of "deliberate practice" (*11*) for the development of expertise. The deliberate practice concept encompasses the educational ideas of constructivism and formative assessment. In our case, the deliberate practice takes the form of a series of challenging questions and tasks that require the students to practice physicist-like reasoning and problem solving during class time while provided with frequent feedback.

The design goal was to have the students spend all their time in class engaged in deliberate practice at "thinking scientifically" in the form of making and testing predictions and arguments about the relevant topics, solving problems, and critiquing their own reasoning and that of others. All of the activities are designed to fit together to support this goal, including moving the simple transfer of factual knowledge outside of class as much as possible and creating tasks and feedback that motivate students to become fully engaged. As the students work through these tasks, they receive feedback from fellow students (*12*) and from the instructor. We incorporate multiple "best instructional practices," but we believe the educational benefit does not come primarily from any particular practice but rather from the integration into the overall deliberate practice framework.

This study was carried out in the second term of the first-year physics sequence taken by all undergraduate engineering students at the University of British Columbia. This calculus-based course covers various standard topics in electricity and magnetism. The course enrollment was 850 students, who were divided among three sections. Each section had 3 hours of lecture per week. The lectures were held in a large theater-style lecture hall with fixed chairs behind benches grouping up to five students. The students also had weekly homework assignments, instructional laboratories, and tutorials and recitations where they solved problems; this work was graded. There were two midterm exams and a final exam. All course components were common across all three sections, except for the lectures, which were prepared and given independently by three different instructors.

During week 12, we studied two sections whose instructors agreed to participate. For the 11 weeks preceding the study, both sections were taught in a similar manner by two instructors (A and B), both with above average student teaching evaluations and many years experience teaching this course and many others. Both instructors lectured using PowerPoint slides to present content and example problems and also showed demonstrations. Meanwhile, the students took notes. "Clicker" (or "personal response system") questions (average 1.5 per class, range 0 to 5) were used for summative evaluation (which was characterized by individual testing without discussion or follow-up other than a summary of the correct answers). Students were given participation credit for submitting answers.

Before the experiment, a variety of data were collected on the students in the two sections

Table 1. Measures of student perceptions, behaviors, and knowledge.

	Control section	Experimental section
Number of students enrolled	267	271
Mean BEMA score (*13*) (week 11)	47 ± 1%	47 ± 1%
Mean CLASS score (*14*) (start of term) (agreement with physicist)	63 ± 1%	65 ± 1%
Mean midterm 1 score	59 ± 1%	59 ± 1%
Mean midterm 2 score	51 ± 1%	53 ± 1%
Attendance before experiment*	55 ± 3%	57 ± 2%
Attendance during experiment	53 ± 3%	75 ± 5%
Engagement before experiment*	45 ± 5%	45 ± 5%
Engagement during experiment	45 ± 5%	85 ± 5%

*Average value of multiple measurements carried out in a 2-week interval before the experiment. Engagement also varies over location in the classroom; numbers given are spatial and temporal averages.

[1]Carl Wieman Science Education Initiative, University of British Columbia, Vancouver, BC, Canada. [2]Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada.

*On leave from the University of British Columbia and the University of Colorado.

†To whom correspondence should be addressed. E-mail: gilbertwieman@gmail.com

‡This work does not necessarily represent the views of the Office of Science and Technology Policy or the United States government.

(Table 1). Students took two midterm exams (identical across all sections). In week 11, students took the Brief Electricity and Magnetism Assessment (BEMA), which measures conceptual knowledge (*13*). At the start of the term, students took the Colorado Learning Attitudes about Science Survey (CLASS) (*14*), which measures a student's perceptions of physics. During weeks 10 and 11, we measured student attendance and engagement in both sections. Attendance was measured by counting the number of students present, and engagement was measured by four trained observers in each class using the protocol discussed in the supporting online material (SOM) (*15*). The results show that the two sections were indistinguishable (Table 1). This in itself is interesting, because the personalities of the two instructors are rather different, with instructor A (control section) being more animated and intense.

The experimental intervention took place during the 3 hours of lecture in the 12th week. Those classes covered the unit on electromagnetic waves. This unit included standard topics such as plane waves and energy of electromagnetic waves and photons. The control section was taught by instructor A using the same instructional approach as in the previous weeks, except they added instructions to read the relevant chapter in the textbook before class. The experimental section was taught by two instructors who had not previously taught these students. The instructors were the first author of this paper, L.D., assisted by the second author, E.S. Instructor A and L.D. had agreed to make this a learning competition. L.D. and instructor A agreed beforehand what topics and learning objectives would be covered. A multiple-choice test (see SOM) was developed by L.D. and instructor A that they and instructor B agreed was a good measure of the learning objectives and physics content. The test was prepared at the end of week 12. Most of the test questions were clicker questions previously used at another university, often slightly modified. Both sections were told that they would receive a bonus of 3% of the course grade for the combination of participating in clicker questions, taking the test, and (only in the experimental section) turning in group task solutions, with the apportionment of credit across these tasks left unspecified.

In contrast to instructor A, the teaching experience of L.D. and E.S. had been limited to serving as teaching assistants. L.D. was a postdoctoral researcher working in the Carl Wieman (third author of this paper) Science Education Initiative (CWSEI) and had received training in physics education and learning research and methods of effective pedagogy while assisting with the teaching of six courses. E.S. had a typical physics graduate student background except for having taken a seminar course in physics education.

The instructional approach used in the experimental section included elements promoted by CWSEI and its partner initiative at the University of Colorado: preclass reading assignments, preclass reading quizzes, in-class clicker questions with student-student discussion (CQ), small-group active learning tasks (GT), and targeted in-class instructor feedback (IF). Before each of the three 50-min classes, students were assigned a three- or four-page reading, and they completed a short true-false online quiz on the reading. To avoid student resistance, at the beginning of the first class, several minutes were used to explain to students why the material was being taught this way and how research showed that this approach would increase their learning.

A typical schedule for a class was the following: CQ1, 2 min; IF, 4 min; CQ2, 2 min; IF, 4 min; CQ2 (continued), 3 min; IF, 5 min; Revote CQ2, 1 min; CQ3, 3 min; IF, 6 min; GT1, 6 min; IF with a demonstration, 6 min; GT1 (continued), 4 min; and IF, 3 min. The time duration for a question or activity includes the amount of time the students spent discussing the problem and asking numerous questions. There was no formal lecturing; however, guidance and explanations were provided by the instructor throughout the class. The instructor responded to student-generated questions, to results from the clicker responses, and to what the instructor heard by listening in on the student-student discussions. Students' questions commonly expanded upon and extended the material covered by the clicker questions or small-group tasks. The material shown on the slides used in class is given in the SOM, along with some commentary about the design elements and preparation time required.

At the beginning of each class, the students were asked to form groups of two. After a clicker question was shown to the class, the students discussed the question within their groups (which often expanded to three or more students) and submitted their answer using clickers. When the voting was complete, the instructor showed the results and gave feedback. The small-group tasks were questions that required a written response. Students worked in the same groups but submitted individual answers at the end of each class for participation credit. Instructor A observed each of these classes before teaching his own class and chose to use most of the clicker questions developed for the experimental class. However, Instructor A used these only for summative evaluation, as described above.

L.D. and E.S. together designed the clicker questions and small-group tasks. L.D. and E.S. had not taught this class before and were not familiar with the students. Before the first class, they solicited two volunteers enrolled in the course to pilot-test the materials. The volunteers were asked to think aloud as they reasoned through the planned questions and tasks. Results from this testing were used to modify the clicker questions and tasks to reduce misinterpretations and adjust the level of difficulty. This process was repeated before the second class with one volunteer.

During the week of the experiment, engagement and attendance remained unchanged in the control section. In the experimental section, student engagement nearly doubled and attendance increased by 20% (Table 1). The reason for the attendance increase is not known. We hypothesize that of the many students who attended only part of a normal class, more of them were captured by the happenings in the experimental section and decided to stay and to return for the subsequent classes.

The test was administered in both sections in the first class after the completion of the 3-hour unit. The control section had covered the material related to all 12 of the questions on the test. The experimental section covered only 11 of the 12 questions in the allotted time. Two days before the test was given, the students in both sections were reminded of the test and given links to the postings of all the material used in the experimental section: the preclass reading assignments and quizzes; the clicker questions; and the group tasks, along with answers to all of these. The students were encouraged by e-mail and in class to try their best on the test and were told that it would be good practice for the final exam, but their performance on the test did not affect their course grade. Few students in either section finished in less than 15 min, with the average being about 20 min.

The test results are shown in Fig. 1. For the experimental section, 211 students attended class to take the test, whereas 171 did so in the control section. The average scores were 41 ± 1% in the control section and 74 ± 1% in the experimental section. Random guessing would produce a score of 23%, so the students in the experimental section did more than twice as well on this test as those in the control section.

The test score distributions are not normal (Fig. 1). A ceiling effect is apparent in the experi-

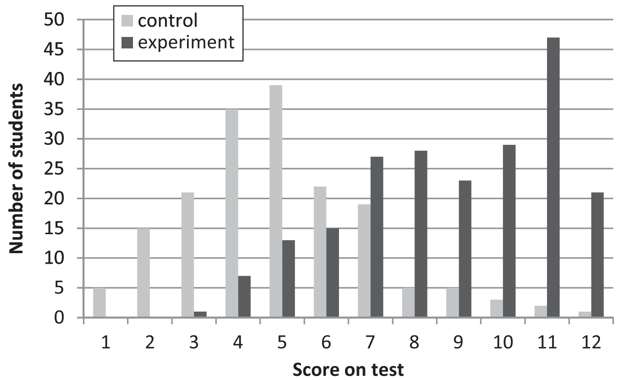

Fig. 1. Histogram of student scores for the two sections.

mental section. The two distributions have little overlap, demonstrating that the differences in learning between the two sections exist for essentially the entire student population. The standard deviation calculated for both sections was about 13%, giving an effect size for the difference between the two sections of 2.5 standard deviations. As reviewed in (*4*), other science and engineering classroom studies report effect sizes less than 1.0. An effect size of 2, obtained with trained personal tutors, is claimed to be the largest observed for any educational intervention (*16*).

This work may obtain larger effect sizes than in this previous work because of the design and implementation that maximized productive engagement. The clicker questions and group tasks were designed not only to require explicit expert reasoning but also to be sufficiently interesting and personally relevant to motivate students to fully engage. Another factor could be that previous work primarily used end-of-term tests, and the results on those tests reflect all the learning that students do inside and outside of class, for example, the learning that takes place while doing homework and studying for exams. In our intervention, the immediate low-stakes test more directly measured the learning achieved from preclass reading and class itself, in the absence of subsequent study.

We are often asked about the possible contributions of the Hawthorne effect, where any change in conditions is said to result in improved performance. As discussed in citations in the SOM, the original Hawthorne plant data actually show no such effect, nor do experiments in educational settings (*17*).

A concern frequently voiced by faculty as they consider adopting active learning approaches is that students might oppose the change (*18*). A week after the completion of the experiment and exam, we gave students in the experimental section an online survey (see SOM); 150 students completed the survey.

For the survey statement "I really enjoyed the interactive teaching technique during the three lectures on E&M waves," 90% of the respondents agreed (47% strongly agreed, 43% agreed) and only 1% disagreed. For the statement "I feel I would have learned more if the whole physics 153 course would have been taught in this highly interactive style." 77% agreed and only 7% disagreed. Thus, this form of instruction was well received by students.

In conclusion, we show that use of deliberate practice teaching strategies can improve both learning and engagement in a large introductory physics course as compared with what was obtained with the lecture method. Our study compares similar students, and teachers with the same learning objectives and the same instructional time and tests. This result is likely to generalize to a variety of postsecondary courses.

References and Notes

1. R. J. Beichner *et al.*, in *Research-Based Reform of University Physics*, E. F. Redish, P. J. Cooney, Eds. (American Association of Physics Teachers, College Park, MD, 2007).
2. C. H. Crouch, E. Mazur, *Am. J. Phys.* **69**, 970 (2001).
3. J. E. Froyd, "White paper on promising practices in undergraduate STEM education" [Commissioned paper for the Evidence on Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics (STEM) Education Project, The National Academies Board on Science Education, 2008]. www7.nationalacademies.org/bose/Froyd_Promising_Practices_CommissionedPaper.pdf
4. J. E. Froyd, "Evidence for the efficacy of student-active learning pedagogies" (Project Kaleidoscope, 2007). www.pkal.org/documents/BibliographyofSALPedagogies.cfm
5. R. R. Hake, *Am. J. Phys.* **66**, 64 (1998).
6. J. K. Knight, W. B. Wood, *Cell Biol. Educ.* **4**, 298 (2005).
7. M. Prince, *J. Eng. Educ.* **93**, 223 (2004).
8. L. Springer, M. E. Stanne, S. S. Donavan, *Rev. Educ. Res.* **69**, 21 (1999).
9. D. Hestenes, M. Wells, G. Swackhamer, *Phys. Teach.* **30**, 141 (1992).
10. L. Deslauriers, C. Wieman, *Phys. Rev. ST Phys. Educ. Res.* **7**, 010101 (2011).
11. K. A. Ericsson, R. Krampe, C. Tesch-Romer, *Psychol. Rev.* **100**, 363 (1993).
12. M. K. Smith *et al.*, *Science* **323**, 122 (2009).
13. L. Ding, R. Chabay, B. Sherwood, R. Beichner, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010105 (2006).
14. W. K. Adams *et al.*, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010101 (2006).
15. Materials and methods are available as supporting material on *Science* Online.
16. B. Bloom, *Educ. Res.* **13**, 4 (1984).
17. R. H. Bauernfeind, C. J. Olson, *Phi Delta Kappan* **55**, 271 (1973).
18. G. K. Allen, J. F. Wedman, L. C. Folk, *Innovative High. Educ.* **26**, 103 (2001).

Acknowledgments: This work was supported by the University of British Columbia through the Carl Wieman Science Education Initiative.

Supporting Online Material

www.sciencemag.org/cgi/content/full/332/6031/862/DC1
Materials and Methods
SOM Text
References

16 December 2010; accepted 5 April 2011
10.1126/science.1201783