Large-scale comparison of science teaching methods sends clear message

Carl E. Wieman[1]

Department of Physics and Graduate School of Education, Stanford University, Stanford, CA 94305

The quality of science, technology, engineering, and mathematics (STEM) education in the United States has long been an area of national concern, but that concern has not resulted in improvement. Recently, there has been a growing sense that an opportunity for progress at the higher education level lies in the extensive research on different teaching methods that have been carried out during the last few decades. Most of this research has been on "active learning methods" and the comparison with the standard lecture method in which students are primarily listening and taking notes. As the number of research studies has grown, it has become increasingly clear to researchers that active learning methods achieve better educational outcomes. The possibilities for improving postsecondary STEM education through more extensive use of these research-based teaching methods were reflected in two important recent reports (1, 2). However, the size and consistency of the benefits of active learning remained unclear. In PNAS, Freeman et al. (3) provide a much more extensive quantitative analysis of the research on active learning in college and university STEM courses than previously existed. It was a massive effort involving the tracking and analyzing of 642 papers spanning many fields and publication venues and a very careful analysis of 225 papers that met their standards for the meta-analysis. The results that emerge from this meta-analysis have important implications for the future of STEM teaching and STEM education research.

In active learning methods, students are spending a significant fraction of the class time on activities that require them to be actively processing and applying information in a variety of ways, such as answering questions using electronic clickers, completing worksheet exercises, and discussing and solving problems with fellow students. The instructor designs the questions and activities and provides follow-up guidance and instruction based on student results and questions. The education research comparing learning from this method with that from the lecture method has usually been carried out by scientists and engineers in the multiple respective disciplines, because the desired learning and the implementation of the teaching methods are quite discipline specific and require substantial disciplinary expertise. Also, good active learning tasks simulate authentic problem solving, and therefore teaching with these methods typically demands more instructor subject expertise than does a lecture.

Probably the most striking result in ref. 3 is that the impact of active learning on educational outcomes is both large and consistent. The authors examined two outcome measures: the failure rate in courses and the performance on tests. They found the average failure rate decreased from 34% with traditional lecturing to 22% with active learning (Fig. 1A), whereas performance on identical or comparable tests increased by nearly half the SD of the test scores (an effect size of 0.47). These benefits of active learning were consistent across all of the different STEM disciplines and different levels of courses (introductory, advanced, majors, and nonmajors) and across different experimental methodologies.

Although the average improvement on tests of all types is substantial, perhaps more notable is the larger improvement on concept inventory (CI) tests, where the effect size is 0.88 (Fig. 1B). CIs are carefully developed tests that probe the differences between how scientists and students think about and use particular scientific concepts. As typically used, CIs also correct for the level of student knowledge at the start of a course and therefore provide a direct measure of the amount learned. Although limited in their scope, CIs are better than instructor-prepared examinations for measuring how well the students have learned to think like scientists.

It is not surprising that the effect size from active learning is larger on CIs. Nearly all techniques labeled as active learning include those features known to be required for the development of expertise (4); in this case, thinking like an expert in the discipline. The active learning methods are designed to
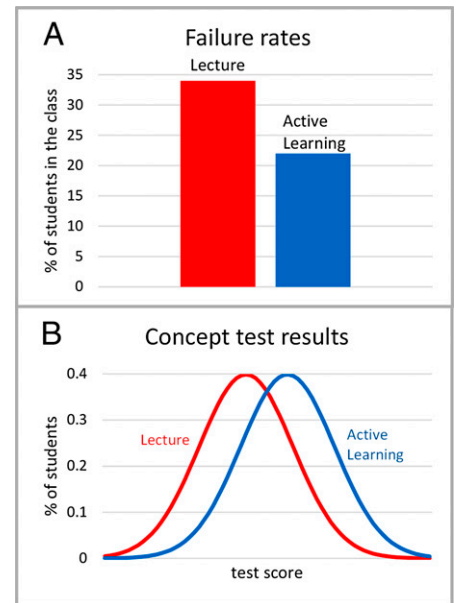

Fig. 1. Comparisons of average results for studies reported in ref. 3. (A) Failure rates for the active learning courses and the lecture courses. (B) Shift in distribution of student scores on concept inventory tests.

have the student working on tasks that simulate an aspect of expert reasoning and/or problem-solving while receiving timely and specific feedback from fellow students and the instructor that guides them on how to improve. These elements of authentic practice and feedback are general requirements for developing expertise at all levels and disciplines (4) and are absent in lectures. Because CI tests are specifically designed to probe expertise developed during a course, they are particularly sensitive to these differences in instructional methods. The relationship between active learning and general requirements for expertise development may also explain the consistency of the benefits across the different disciplines and levels of courses.

The implications of these meta-analysis results for instruction are profound, assuming they are indicative of what could be obtained if active learning methods replaced

the lecture instruction that dominates US postsecondary STEM instruction. With a total annual enrollment in STEM courses of several million, a reduction in average failure rate from 34% to 22% would mean that an enormous number of students who are now failing STEM courses would instead be successfully completing them. The expected gains in learning for all students in STEM courses are equally important.

However, such gains should be considered only the minimum of what is possible. The variations in study results tabulated in ref. 3 are not random fluctuations; they are the result of differences in the active learning methods used. As further research identifies the relative benefits of different active learning methods and the most effective means of implementation, substantially larger impacts can be expected. One promising direction emerging from ref. 3 is that "more is better." The highest impacts are observed in studies where a larger fraction of the class time was devoted to active learning. Those high impact studies would suggest that it is reasonable to aspire to teaching that consistently achieves twice the average improvements reported in ref. 3, e.g., failure rates of only ~10% and increases in learning of 1–1.5 SDs. Such improvements in STEM educational outcomes would have major national implications.

This meta-analysis makes a powerful case that any college or university that is teaching its STEM courses by traditional lectures is providing an inferior education to its students. One hopes that it will inspire administrators to start paying attention to the teaching methods being used in their classrooms—monitoring them and establishing accountability for using active learning methods, something that is currently not done.

The results in this meta-analysis also have implications for science education research. First, the consistency of results across different research methodologies (quasi random vs. random, failure rates on identical vs. non-identical examinations, different types of

Freeman et al. argue that it is no longer appropriate to use lecture teaching as the comparison standard.

instructor comparisons, etc.) addresses concerns that have been raised about some methodologies. This consistency is different from the situation in K-12 education research, where observed effect sizes are smaller and more variable across research designs. A likely reason for this difference is that the K-12 context (including instructor subject mastery) and the student population are far less controlled and consistent. This means there are many more confounding variables that can affect results.

A particularly interesting finding is that the educational results were similar for the different ways of selecting instructors for comparison. Comparative results between lecture and active learning were the same for one instructor using the two different methods or independent instructors using

different methods. Thus, there was no indication that the relative effectiveness of the different teaching methods is instructor dependent.

Perhaps the most significant implication for future postsecondary STEM education research is that this meta-analysis will force researchers to think more deeply about their goals. Freeman et al. argue that it is no longer appropriate to use lecture teaching as the comparison standard, and instead, research should compare different active learning methods, because there is such overwhelming evidence that the lecture is substantially less effective. This makes both ethical and scientific sense. If a new antibiotic is being tested for effectiveness, its effectiveness at curing patients is compared with the best current antibiotics and not with treatment by bloodletting. However, in undergraduate STEM education, we have the curious situation that, although more effective teaching methods have been overwhelmingly demonstrated, most STEM courses are still taught by lectures—the pedagogical equivalent of bloodletting. Should the goals of STEM education research be to find more effective ways for students to learn or to provide additional evidence to convince faculty and institutions to change how they are teaching? Freeman et al. do not answer that question, but their findings make it clear that it cannot be avoided.

1 Singer S, Nielsen N, Schweingruber H (2012) *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering* (National Academies Press, Washington, DC).
2 PCAST (2012) Report to the President: Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Available at http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf. Accessed April 20, 2013.

3 Freeman S, et al. (2014) Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci USA* 111:8410–8415.
4 Ericsson A-K (2006) The influence of experience and deliberate practice on the development of superior expert performance. *The Cambridge Handbook of Expertise and Expert Performance*, eds Ericsson A-K, Charness N, Feltovich P-J, Hoffman R-R (Cambridge Univ Press, Cambridge, UK), pp 683–703.