Abstract Title Page

Not included in page count.

Title: Effective Secondary Science Programs: A Best-Evidence Synthesis

Authors and Affiliations: Alan Cheung, The Chinese University of Hong Kong, alancheung@cuhk.edu.hk; Robert E. Slavin, Johns Hopkins University and University of York, rslavin@jhu.edu; Cynthia Lake, Johns Hopkins University, clake5@jhu.edu; and Elizabeth Kim, Johns Hopkins University, lizkim@jhu.edu

Abstract Body.

Background / Context:

Despite widespread recognition among policy makers, educational leaders, and the nation as a whole of the importance of science, engineering, and technology as drivers of the future of our country and society, the science achievement of America's students is mediocre at best, in comparison to that of our international peers. On the 2012 PISA tests in science, U.S. 15-year-olds ranked 28th, slightly below the average of the 65 participating countries. On the 2011 National Assessment of Educational Progress (NCES, 2012) science assessment, 68% of eighth graders scored below "proficient."

In recent years, there has been growing consensus about the goals of science education in elementary and secondary schools, and an acceleration in the use of rigorous quantitative methods to evaluate innovative science methods (Marx, 2012; Penuel & Fishman, 2012). Experimental studies of science teaching innovations have long been done, but most have been brief and artificial experiments, have lacked control groups, have used measures closely aligned to experimental but not control treatments, or have otherwise allowed for the possibility that outcomes will greatly overstate the impacts teachers will actually observe in practical applications.

As the number of rigorous experiments evaluating various science approaches increases, it becomes both necessary and possible to use accepted meta-analytic methods to summarize the findings of studies meeting clearly specified criteria. It is possible to learn from any study, but findings from studies that compare the gains of experimental and control students on fair and valid measures are likely to be of particular value in building a science of educational practice, and to offer solid evidence to education leaders and policy makers about the types of interventions most likely to increase student achievement.

Slavin, Lake, Hanley, & Thurston (2014) reviewed rigorous quantitative research on science programs at the elementary level. The elementary review emphasized three types of programs: inquiry-oriented programs without science kits, inquiry-oriented programs with science kits, and technology programs. Twenty-three studies met these criteria. Programs that used science kits did not show positive outcomes on science achievement measures (weighted ES=+0.02 in 7 studies). However, inquiry-based programs that emphasized professional development but not kits did show significant positive outcomes (weighted ES=+0.36 in 10 studies). The largest effect sizes came from a small number of studies of technology approaches integrating video and computer resources with teaching (weighted ES=+0.42 in 6 studies). The review concluded that science programs focused on improving teachers' classroom instruction, such as cooperative learning and science-reading integration, as well as programs making innovative uses of technology closely integrated with teachers' instruction, are promising avenues for improving science teaching and learning.

Purpose / Objective / Research Question / Focus of Study:

The intention of this review is to place all types of programs designed to enhance the science achievement of middle and high school students on a common scale, to provide educators with meaningful, unbiased information that they can use to select programs most likely to make a difference for their students' learning. In addition, the review is intended to look broadly for common factors that might underlie effective practices across programs and program types, and to inform an overarching understanding of effective instruction in secondary science.

This synthesis also seeks to identify common characteristics of programs likely to make a difference in student science achievement. It includes all types of approaches to science instruction, grouping them in four categories. *Instructional process* approaches were ones that provided substantial training and coaching to teachers in specific approaches to inquiry-oriented science teaching, such as cooperative learning and metacognitive strategy instruction. *Technology programs* are ones that make extensive use of computers to enhance science learning. *Science kits* are programs that provide teachers with inquiry-oriented kits facilitating hands-on experiments, as well as extensive professional development to use the kits. *Textbook programs* provide innovative or standards-based content, but far less PD than instructional process approaches and minimal or no use of technology.

Setting: The studies took place in middle and high schools.

Population / Participants / Subjects:
A total of 21 qualifying studies based on over 31,000 students in grades 6-12 met the inclusion criteria. Ten studies were quasi-experiments (including randomized quasi-experiments) and 11 were randomized studies. Findings were reported in 16 published articles and five in unpublished documents such as dissertations and technical reports. Two of the studies were published in the 1990s, 8 in the 2000s, and 11 in the 2010s. In terms of sample size, eight were small-scale studies (N<250) and 13 were large-scale studies (N≥250). These studies covered a wide range of science subjects, including physics, chemistry, biology, and general science. Eleven of these studies examined the program impacts on middle or junior high school students and ten on senior high school students.

Intervention / Program / Practice:
Criteria for inclusion of studies in this review were as follows:
1. Students had to be in middle or high schools
2. Minimum program duration was 12 weeks.
3. The studies evaluated programs and practices used in secondary science, and were published in 1990 or later. Studies could have taken place in any country, but the reports had to be available in English.
4. The studies took place in middle and high schools.
5. The studies compared students taught in classes using a given science program or practice with those in control classes using an alternative program or standard methods.
6. The program or practice had to be one that could, in principle, be used in ordinary science classes (i.e., it did not depend on conditions unique to the experiment).
7. Random assignment or matching with appropriate adjustments for any pretest differences (e.g., analyses of covariance) had to be used. Random assignment could be at the level of individuals or clusters (e.g., schools or classes). If random assignment was at the cluster level but there were too few clusters for analysis accounting for clustering, this is termed a “randomized quasi-experiment” (Slavin, 2008) and categorized as a matched study. Studies without control groups, such as pre-post comparisons and comparisons to “expected” scores, were excluded.
8. Pretest data had to be provided, and there could be no indications of initial inequality.
9. The dependent measures included quantitative measures of science performance. Experimenter-made measures were accepted if they covered content taught in control as well as

experimental groups, but measures of science objectives inherent to the program (and unlikely to be emphasized in control groups) were excluded.
10. A minimum study duration of 12 weeks was required.
11. Studies had to have at least two teachers, two schools, and 15 students in each treatment group. This criterion reduced the risk of confounding teacher, class, or school effects with treatment effects.

Research Design: A broad literature search was carried out in an attempt to locate every study that could possibly meet the inclusion requirements. A meta-analysis calculated outcomes in effect sizes, and these were pooled within well-justified categories.

Data Collection and Analysis: Electronic searches were made of educational using different combinations of key words and the years 1990-2015. Results were narrowed by subject area, and we also conducted searches by program name. Web-based repositories and education publishers' websites were examined. We contacted producers and developers of secondary science programs to check whether they knew of studies we might have missed. Citations from other reviews of science programs, as well as studies cited in primary research, were obtained and investigated. We searched tables of contents of key journals. Studies that met an initial screen were independently read and coded by at least two researchers. Any disagreements in coding were resolved by discussion, and additional researchers were asked to read any articles on which there remained disagreements.

Findings / Results:
Twenty-one studies met inclusion criteria including use of randomized or matched assignment to conditions, measures that assess content emphasized equally in experimental and control groups, and a duration of at least 12 weeks. Programs fell into four categories. Instructional process programs (ES=+0.24) and technology programs (ES=+0.47) had positive sample-size weighted mean effect sizes, while use of science kits (ES=+0.05) and innovative textbooks (ES=+0.10) had much lower effects.

Instructional process programs. Instructional process science approaches are ones that provide teachers with extensive training and/or classroom coaching in specific teaching methods, such as cooperative learning, use of metacognitive strategies, and project-based learning. Methods of this kind that have been evaluated in qualifying studies have well-defined models of what classroom science teaching should look like and use extensive professional development to help teachers adopt and implement the innovative models. Instructional process models vary substantially, so they should not be seen as a consistent strategy. What they share is a theory of action emphasizing improving student science learning by providing extensive professional development designed to change teachers' behaviors, rather than focusing primarily on innovative materials, texts, or technology with limited PD. Seven programs in the instructional process category each had one qualifying evaluation. The weighted mean effect size was +0.24

Technology programs, of course, have in common the extensive use of digital devices. Their theories of action emphasize the power of digital media to provide material appropriate to students' needs and to integrate visual and text elements of science concepts. Remarkably, none of the technology approaches identified for this review used the computer-assisted instruction

(CAI) strategies that have dominated the use of technology in math and reading for many years, which research has not generally supported (see Cheung & Slavin, 2013 a, b; Slavin, Lake, & Groff, 2009). In science, at the elementary as well as the secondary levels, technology is more often used to illustrate science concepts, simulate real-world processes, and support the teacher's instruction, rather than operating separately.

Only five programs, each with just one evaluation, met the inclusion requirements, but these studies had by far the most positive average effect size among the four categories of programs in the review. The weighted mean was +0.47, even higher than the mean of +0.37 reported for the six qualifying technology studies in elementary science by Slavin et al. (2014).

The focus of the technological applications in secondary science mostly included strategies designed to help students visualize science concepts, and to connect to resources beyond the classroom. Four of the five qualifying studies were small (n<250) and all five used matched rather than randomized assignment to conditions, so these findings are far from conclusive.

Science Kits. Just one study evaluated a program in which teachers were given science kits to help students do experiments and other inquiry-oriented activities. Unlike textbook programs, such approaches provide a great deal of professional development to teachers, and unlike professional development programs they provide specific, well-developed classroom materials that engage students in hands-on experiments and explorations. The effect size for the single study was +0.05.

Textbooks. Textbook innovations represent an approach to science education reform emphasizing the content of courses. The theory of action behind textbook approaches assumes that standards-based content or other features of texts will improve student science outcomes. Professional development is invariably provided to teachers to help them use new textbooks, but on the order of two to five hours at most, in contrast to the five days or more typical of professional development approaches. Also, in textbook methods the innovation is in the content rather than the teacher's instructional methods, which is why less professional development is provided. Eight studies of five textbook innovations found very small impacts (weighted mean effect size =+0.10) for innovative textbooks.

Conclusions

Outcomes support the use of programs with a strong focus on professional development, technology, and support for teaching, rather than materials-focused innovations. The types of programs that make a difference in student outcomes are those that help teachers teach more effective lessons: technology designed primarily to help students visualize science concepts, and instructional process models that provide teachers with extensive professional development to help them apply strategies such as cooperative learning, use of metacognitive skills, and science-literacy integration.

In contrast, approaches that attempt to improve science learning primarily through improving textbooks or providing teachers with kits to facilitate experiments have been less successful in rigorous evaluations.

What these findings imply is that teaching, not materials, is the core of the science classroom, and that investing in specific technologies and professional development designed to enhance the effectiveness of teaching is the best way forward in science education.

Appendices

Appendix A. References

Cheung, A., & Slavin, R. E. (2013a). Effects of educational technology applications on reading outcomes for struggling readers: A best-evidence synthesis. *Reading Research Quarterly, 48*, 277-299.

Cheung, A., & Slavin, R. E. (2013b). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88-113.

Marx, R. W. (2012). Large-scale interventions in science education: The road to utopia? *Journal of Research on ScienceTeaching, 49,* 420–427.

National Center for Education Statistics (2012). *The Nation's Report Card: Science 2011* (NCES 2012–465). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Penuel, W. R., & Fishman, B. J. (2012). Large-scale intervention research we can use. *Journal of Research in Science Teaching, 49,* 281-304.

Slavin, R. E., Lake, C., Hanley, P., & Thurston, A. (2014). Experimental evaluations of elementary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching, 51,* 870-901.

Slavin, R.E., Lake, C., & Groff, C. (2009a). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research, 79,* 839-911.