

Abstract Title Page
Not included in page count.

Title: Communicating to policymakers: what to do with ambiguous results from three impact studies of the same education program

Authors and Affiliations: Ardice Hartry, UC Berkeley

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

This poster considers one challenge from recent emphasis on rigorous research for decision-making in education: the interpretation of non-replicated findings. Under Clinton and Bush, the federal government called for and funded highly rigorous research. The “gold-standard” for research was randomized-control trials (RCTs), which supposedly offered clear answers about program efficacy. Such clear answers were meant to be used by stakeholders in making decisions about adopting programs.

During 2005—2008, we conducted an RCT meant to measure the impact of a reading program in the afterschool setting, with the goal of comparing the effectiveness of that program with regular school offerings. The study covered three school years, with slight variations in both the delivery of the intervention and the design of the research, resulting in three essentially discrete studies conducted within a single school district. We hoped that by replicating the study in one context, we would be able to draw conclusions about the effectiveness of the reading program. This poster, therefore, addresses the following questions:

How may contradictory findings be used in program design and implementation?

How may the conversation among practitioners, policymakers and researchers be expanded to acknowledge both consistency and contradictions in robust research studies?

How may we better communicate what contradictory findings suggest about practice and program development to practitioners?

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

The question of what works well in academic afterschool programs is important because many districts face critical choices when providing supplemental educational services (SES) for schools that do not make adequate yearly progress (AYP). SES programs (often equated with tutoring) can be expensive and challenging to implement. Districts are searching for proven, affordable, and replicable interventions to help their students improve and, through meeting their students’ needs, make AYP.

During the 2005–06 school year (Phase I), the author led a randomized control trial (RCT) of a structured reading intervention in afterschool programs, funded by a national foundation. A second phase of the study, conducted in 2006-2008, was designed to replicate that research, comparing the reading program with another academically focused afterschool program. The reading program was introduced to four new elementary schools in the same district where Phase I was conducted. The purpose of Phase II was to test the validity and reliability of the Phase I findings. Phase II was commissioned by the Office of Elementary and Secondary Education, U. S. Department of Education, and funded under contract through the National Partnership for Quality Afterschool Learning at the Southwest Educational Development Laboratory.

Setting:

Description of the research location.

The study took place in a mid-sized, urban fringe school district on the east coast. This school district had an afterschool program that included some academic curricula as well as enrichment activities. Across the three years of the study, a total of six elementary schools, all from this same district, participated in the program.

Population / Participants / Subjects:

Description of the participants in the study: who, how many, key features, or characteristics.

The study involved approximately 800 students over the three-year period. These students were enrolled in 4th, 5th, and 6th grades. The population of students was approximately 35% white, 50% African-American, 10% Latino, and 5% other, including mixed-race. Over 90% of students were enrolled in the federal free-reduced lunch program, a marker commonly used in education to indicate socioeconomic status.

Intervention / Program / Practice:

Description of the intervention, program, or practice, including details of administration and duration.

The intervention was a structured reading program, designed originally for in-school but modified for the afterschool setting. Students in the treatment group received one hour of the program each day in afterschool, which ran from September through May. Regular classroom teachers function as instructors in the afterschool setting, and they received professional development in how to use the reading program. The reading program consists of three main components: group direct instruction, time using an instructional software program, and student-directed independent reading. At the end of each day, the group is brought back together for a whole-group wrap-up and reflection session, lasting about 10 minutes. The program features many characteristics of high-quality afterschool programs, which have been identified by the National Partnership for Quality Afterschool Learning (Birmingham 2005).

Research Design:

Description of the research design.

The study used random assignment experimental design. Our study population consisted of all 4th, 5th, and 6th grade students who were enrolled in the afterschool program at four school sites. Students were randomly assigned into one of two groups. One group, consisting of half of the students, received one hour of homework and enrichment, and one hour of other activities, some of which were organized and had a literacy component, others of which were less structured and may or may not have contained literacy activities. The other group received one hour of homework and enrichment and one hour of a structured reading program (the intervention of interest). Because this program had not been designed for use in the afterschool setting, however, it was modified from its original design, in order to fit into the afterschool schedule. The ways in which the intervention was modified changed from year to year. We intended to compare student outcomes in reading achievement and attitudes towards reading between treatment and control groups. Because it was a replication study, we hoped that we would find similar outcomes across all three years. This design was the heart of the evaluation.

In addition, the design included an implementation or fidelity study. Using data from site visits (including interviews and classroom observations) ongoing, rigorous evaluation of the extent and fidelity of the implementation of the intervention was conducted to ensure that students received the treatment as designed. The degree of implementation and the dosage received by each student, and delivered at each school, was included as an explanatory variable in the analysis.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.

The study looked at two major categories of student outcomes: student achievement in literacy and student attitudes towards reading. Student achievement consisted of standardized test scores, assessments of oral reading fluency, and book title recognition tests.

To ascertain levels of student engagement, the study collected data on student attendance at the afterschool program and administered surveys to students about their attitudes towards reading and school.

Our analysis addressed the following research questions:

- Did participation in the reading program improve reading skills more than typical afterschool activities?
- Did the reading program work better to improve outcomes related to academic achievement, such as afterschool attendance and attitudes towards reading?
- Did the reading program work equally well for different subgroups of students, including children who varied in ethnicity, disability status, grade level, reading abilities, and gender?

Differences in assessment outcomes between the treatment and control groups were estimated using ordinary least squares regression. The outcome variable of interest was the student score on the achievement test administered at the end of the year. Covariates includes pre-test score administered prior to random assignment, dummy variables indicating inclusion in either treatment or control, and dummy variables that indicated blocking factors by grade and school. All estimates are based on intention to treat.

Findings / Results:

Description of the main findings with specific details.

The results of our analyses, however, were ambiguous. In the first year (Phase I), small indications of success were found. Out of the three achievement measures we used, we found only differences on one, and only at the fourth grade. In the fifth grade, small differences were found but these were not significant. In sixth grade, differences were again not significant – and they were negative. We also found no differences between treatment and control on the attitudinal measures.

In the second year (Phase II), much larger differences between treatment and control students were detected, but on different measures than in year one. First, on the standardized reaching achievement measure, we found very significant differences between treatment and control, and in all three grades – but particularly in fourth grade. As in Year 1, we found no differences on the attitudinal measures.

In the third year, virtually no differences were found, on any measures, between any groups. Interestingly, the size of the difference between the scores of the treatment group and the scores of the control groups were as large in our study as they were in a much larger study that found significant differences. Because of the smaller number of students in our study, however, these differences did not appear significant. Once again, we found no differences on the student survey of attitudes. In summary, the student achievement measures across the three years were mixed in Year 1, positive in Year 2, and neutral in Year 3.

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

So how do we make sense of these data? We feel it is important to begin with a description of implementation – of both the intervention and the research – in detail, and so, in the case of our research, the first article we published was about the implementation itself. Implementation was not perfect, but it met the requirements for high levels of implementation, especially considering the fact that it was altered to fit the afterschool program setting. But, because it was altered for the afterschool setting, the way in which it was altered varied from year to year, as the developer sought to improve the delivery based upon formative evidence from the implementation study.

So what decisions should be made on the basis of these studies? If you were the Superintendent of a school district, or a teacher sitting on a reading intervention committee, or a School Board member, or bureaucrat at the state office of education, would you recommend this program for adoption? A policymaker or stakeholder would be able to say, with some degree of support, that implementing this program is possible, and it is potentially effective if implemented with the correct group of students. The effects, however, are small, and it is possible that a policymaker could also conclude that it is not worth the cost of the program to adopt it without more certainty that it will make a big difference in the reading achievement of students.

Studies should collect and publish detailed data on implementation of both the treatment and of the research itself. While the problems with not presenting implementation data have become increasingly well recognized in the last five years, implementation data are often not presented in any consistent fashion, or in great detail. It may not always be possible to present these data within the confines of a single journal article, but surely multiple journal articles, or large-scale publicly available reports, can address this problem. But understanding the context of a study, and the details of the implementation of the program, can help policymakers and stakeholders make important decisions about the likelihood that the program will be effective in their schools, with their students. Without both quantitative data about outcomes and qualitative data about implementation, educators are making decisions with only half the information they need. Mixed method approaches, especially within the confines of experimental research, are vital to improving education.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Birmingham, Jennifer, Pechman, Ellen M., Russell, Christina A., and Mielke, Monica. (2005). Shared Features of High-Performing After-School Programs: A Follow-Up to the TASC Evaluation. Policy Studies Associates, Inc.

Cunningham, Anne & Stanovich, Keith (1991). Tracking the Unique Effects of Print Exposure in Children: Associations With Vocabulary, General Knowledge, and Spelling. *Journal of Educational Psychology*, 83(2), 264–274.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5, 239–256.

Gambrell, L. B., Palmer, B. M., Codling, R. M., & Mazzone, S. A. (1996, April). Assessing motivation to read. *The Reading Teacher*, 49(7), 518–533.

Good, R. H., & Kaminski, R. A. (Eds.). (2003). *Dynamic indicators of basic early literacy skills* (6th ed., revised 07/02/03). Eugene, OR: Institute for the Development of Educational Achievement.

Hasbrouck, J., & Tindal, G. (2005). *Oral reading fluency: 90 years of measurement*. Technical Report #33. Eugene, OR: College of Education, University of Oregon.

Harcourt Assessment, Inc. 2004. *Technical Data Report*, Stanford Achievement Test Series Tenth Edition.

Hill, C., Bloom, H., Black, A., & Lipsey, M. (July 2007). Empirical Benchmarks for Interpreting Effect Sizes in Research. MDRC Working Paper on Research Methodology. New York: MDRC.

McKenna, M. C., & Kear, D. J. (1990, May). Measuring attitudes toward reading: A new tool for teachers. *The Reading Teacher*, 626–639.

McKenna, M. C., & Kear, D. J., & Ellsworth, R.A. (1995). Children's Attitudes Toward Reading: A National Survey. *Reading Research Quarterly* (30/4), 934–955.

Appendix B. Tables and Figures
Not included in page count.