



[Buy this issue](#)

- [Read Abstract](#)

February 2015 | Volume 72 | Number 5  
**Improving Schools: What Works?** Pages 10-14

## Trust, But Verify

*David C. Berliner and Gene V Glass*

**A program that sparkles in Maine may fizzle out in Montana. If we can't depend on program transferability, what's a leader to do?**

School improvement programs that work in some places don't work in others. School improvement programs that work with some students don't work with others. Programs that appear to have positive effects in the hands of some teachers don't work for other teachers. If this were not the reality of school improvement, we would have found and implemented excellent programs for every state, district, and classroom in the United States by now.

But we haven't. Instead, we're continually puzzled as we search for high-quality education programs for rural white students, for urban black students, and for English language learners from hundreds of nations. We also have problems educating the privileged youth of America's upper-class communities, who sometimes seem not to understand the responsibilities that come with our democracy and the role each of us must play to maintain the commons. The education of children who suffer from "affluenza" (Fernandez & Schwartz, 2013) is as disappointing to many educators as is the slow progress of America's poor students.

Perhaps it's time to lay aside the belief that what works in one setting with one teacher at one time is highly likely to work in another setting with another teacher at another time. Education, says our colleague Lenay Dunn (Berliner, Glass, & Associates, 2014) is a complex, intricate endeavor that entails inputs we can't control (for instance, family wealth, parents' education, community support, and special needs of children); variables we can't easily identify or measure (such as competing school and district initiatives, classroom culture, peer influence, teacher beliefs, and principal leadership); and outputs we can neither predict nor easily measure (such as resilience, grit, practical intelligence, social intelligence, and creativity). The complex nature of education limits the ability of schools and districts to duplicate programs that have performed well elsewhere.

We might despair in the face of this reality. Or we might, instead, feel privileged that we work in a field that is more complex, and thus more challenging, than physics or rocket science. The late, great economist Kenneth Boulding once remarked that if physical systems were as complex as social systems, we would creep hesitantly out of bed each morning, not knowing whether we were about to crash to the floor or float to the ceiling. Educators face the challenge of these unpredictable social systems every day.

### Three Obstacles to Transfer

Education is simply too complex an endeavor to allow us the kind of certainty that characterizes the natural sciences, where a finding is a finding is a finding, where whatever was found to hold true in Rio de Janeiro is also true for Los Angeles, and rural

Mississippi, and on rainy as well as sunny days.

Contexts matter in the social sciences. Because of their complexity, we may never understand how to assess the power of all the interacting variables that make up a context, and thus we may never be able to predict when and where a program will and will not work. But it's more than the complexities of context that limit our confidence in a program's transferability to a different setting. Three additional problems make it difficult to transfer programs that appear to work to a new and different setting.

### **The Problem with Findings**

First is the problem of estimating the power of the program that we want to import to our school or district. How strong were the original findings? Were the effects strong enough to suggest that we ought to try it elsewhere? Many reports of a successful program or activity present their results as "statistically significant." But that doesn't mean much because statistical significance is primarily a reflection of sample size. A pill that works for only one person out of 50 can produce a statistically significant result in a huge clinical trial.

Interpreting data also requires knowledge of whether random assignment occurred and whether the investigators were the same people who developed the program under study. It's better to have data about a program's effects presented as an effect size, which helps us decide whether the program's effect, despite all the complications in the study's design, is potentially large enough to be worth pursuing in terms of time, money, and personnel costs.

But even if the overall effect of a program was impressive, the conditions under which the program did *not* work are rarely discussed and are not well understood. The famous Tennessee class-size study (Mosteller, 1995), the STAR study, showed impressive overall benefits of smaller classes. Since that study was published, many have argued that major reductions in class size for poor children are likely to have lasting effects on the children's lives.

But Konstantopoulos (2011) looked *within* the overall data and noted that

*results revealed that a large proportion of the school-specific small class effects are positive, while a smaller proportion of the estimates are negative. Although students benefit considerably from being in small classes in many schools, in other schools being in small classes is either not beneficial or is a disadvantage. Small class effects were inconsistent and varied significantly across schools in all grades. (p. 71)*

This is no different a result from what we find in pharmacological studies. A drug may turn out to have an overall average positive effect, and thus is approved by the Food and Drug Administration. Forgotten in the rush to bring the drug to market are the data that show it didn't work for many in the sample, it harmed some, and among those who showed positive effects were many people who responded because of placebo effects. Pharmacological research is closer to education research than research in the natural sciences is.

Just as human biological systems vary, and drugs work with some patients and not with others, school and class contexts vary a great deal. Programs like class-size reduction are fine candidates for improving the progress of poor students and the working conditions of teachers, but they may not always work as we hope.

Konstantopoulos's insights into the effects of the class-size study are similar to the advertisements for medicines one hears on television. You hear about how wonderful a drug is—just before the fast talk begins informing you that it may produce blood clots, susceptibility to tuberculosis, increased heart problems, and the like. We eventually learn that overall success is invariably accompanied by many noneffects and quite a few failures.

But few researchers, and even fewer promoters of programs, do the high-quality research that would reveal noneffects, or negative effects for some children, when a given program is in the hands of some teachers and in certain schools. Education research doesn't provide us with such answers.

### **The Problem with Replicability**

The gold standard of research is often said to be the randomized clinical trial. But we don't think so. The real standard is a replication of effects by authors who neither produced the original study nor designed the original program.

Try not to get scared, but in medicine, one major study suggested that only 44 percent of the replications of medical research produced supportive data (Makel & Plucker, 2014). Unsuccessful replications most often occurred when the sample size in the original study was small and when randomization was not employed.

These are precisely the conditions that describe a great deal of education research. But we don't have a non-confirmation problem in education research, as does medicine, because we have an even more serious problem: We don't even do replication research! The replication rate for research in our top journals, at well under 1 percent, is frighteningly low. The lack of replications, of course, makes it harder to be confident that a program that works in one location will work in another.

### **The Problem with Fading Effects**

As teachers change, as student characteristics change, as assessment instruments change, and as school cultures change, a program that seemed successful a few years back may no longer work as it did. Programs need to be monitored for efficacy over time, just as medicines do. Also, ideas that are key to the program of interest may already be in place among the students we want to help, and so bringing the new program in shows little or no effect.

Lemons, Fuchs, Gilbert, and Fuchs (2014) examined five randomized studies of a supplemental peer-mediated kindergarten reading program involving more than 2,500 students across nine years. They found a dramatic increase in the performance of the control-group students over time. Obviously, if the control groups are doing better on the measures used to evaluate a program's efficacy, it's harder for the program to show an effect in a new district or school. The students in the control groups somehow were getting better instruction over time, so the power of the peer-mediated reading program to show its effects got weaker and weaker. We rarely have nuanced or complete data about the students we want to help when we bring in a new program, and this lack of understanding may weaken the effects we finally see.

The whole idea of "bringing programs to scale" (that is, moving a program from a few schools to many) is also a problem. Control of the contextual complexity in a few classes, or in a school or two, is a lot easier than control of the myriad contextual variables affecting programs in entire districts or states.

## Realistically Optimistic

So things don't always work as expected. What are school leaders to do? The best they can! Some data are probably better than no data, if collected honestly by individuals who aren't out to make a lot of money by pushing a program.

So look at the data. But overselling an idea or program in your own district is a mistake. You'll need to try it out, probably adapt it to local circumstances, and then it still may not work as intended. But it might.

A realistic view of the difficulties that lie in the path to school improvement must not lead to despair. As professionals, we're expected to seek better ways of educating children. Trying out programs that have been successful elsewhere, designing new programs that fit local circumstances, and attempting to implement what sound like good ideas are characteristic of exemplary leadership.

Three considerations will increase the chances that experimentation will lead to improvement. One is having teacher buy-in. Not much works well if teachers have things imposed on them that they don't believe in. Second, don't implement several new programs and ideas simultaneously. Teachers often suffer from overload when new administrators, or state and federal bureaucrats, set out to change too many things too quickly. Finally, make sure new programs and ideas undergo a formative evaluation to find out how things work and how they might be improved. This might entail asking a local evaluator or colleagues from a different school to help with formative and summative assessments of a program.

In 1987, at the signing of a treaty with the Soviet Union, President Reagan remarked, "Trust, but verify." His advice is our advice: Trust that your colleagues across the United States and around the world have found some good ideas for school improvement that work for them. But verify that their thinking will work for you, too.

## Ideas That (May) Travel Well

Here are a few pet ideas that we've seen work in one place or another that might offer alternative approaches to school improvement:

- Stop looking for answers to local problems in Scandinavia or Asia. The United States is neither Finland nor Singapore, and it's a lot more complex than either.
- Redraw school attendance areas to achieve socioeconomic balance, and support high-quality early childhood education in those areas.
- Recognize that teachers work in teams and evaluate them accordingly. Make sure the evaluation system has no consequences for teachers associated with student test scores but does include multiple classroom observations and an evaluation of classroom artifacts—tests, papers, projects, and the like.
- Eliminate tracking in grades K–6, and eliminate grade retention ("flunking") completely.
- Make sure that no school day for students starts earlier than 8:30 a.m.
- Provide libraries staffed with librarians and counseling offices staffed with enough counselors that they can know students personally.
- If you don't like your reading scores, find ways to have students read more, and forget most other systems that claim to improve reading.

## References

Berliner, D. C., Glass, G. V., & Associates. (2014). *50 myths and lies that threaten America's public schools*. New York: Teachers College Press.

Fernandez, M., & Schwartz, J. (2013, December 13). Teenager's sentence in fatal drunken-driving case stirs "affluenza" debate. *New York Times*. Retrieved from [www.nytimes.com/2013/12/14/us/teenagers-sentence-in-fatal-drunken-driving-case-stirs-affluenza-debate.html](http://www.nytimes.com/2013/12/14/us/teenagers-sentence-in-fatal-drunken-driving-case-stirs-affluenza-debate.html)

Konstantopoulos, S. (2011). How consistent are class size effects? *Evaluation Review*, 35(1), 71–92.

Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher*, 43(5), 242–252.

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43(6), 304–316.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *Future of Children*, 5(2), 113–127.

[David C. Berliner](#) is an educational psychologist and Regents' Professor Emeritus of Education at Arizona State University. [Gene V Glass](#) Regents' Professor Emeritus at Arizona State University, senior researcher at the National Education Policy Center, and a research professor in the School of Education at the University of Colorado, Boulder. They are coauthors of *50 Myths and Lies That Threaten America's Public Schools: The Real Crisis in Education* (Teachers College Press, 2014).

## KEYWORDS

Click on keywords to see similar products:

[school improvement and reform](#), [Audience: All Educators](#), [Education Level: All Grade Levels](#)

Copyright © 2015 by ASCD

## Requesting Permission

- For **photocopy, electronic and online access, and republication requests**, go to the [Copyright Clearance Center](#). Enter the periodical title within the "**Get Permission**" search field.
- To **translate** this article, contact [permissions@ascd.org](mailto:permissions@ascd.org)