EERJ | European Educational Research Journal

ECER GHENT KEYNOTE

Evidence-Based Reform in Education: what will it take?[1]

ROBERT E. SLAVIN
University of York, United Kingdom, and Johns Hopkins University, USA

During the twentieth century, medicine, agriculture, technology, and other fields embraced a simple but powerful idea: Use what works. They began to require that innovative medicines, seeds, and machines be put to the test before being widely adopted. The result was revolutionary progress in each of these fields, which continues today. Evidence-based reform in any area does not just protect the public from ineffective innovations; it also creates a dynamic of progressive improvement, in which many researchers and developers are working to replace today's best solutions with something even more effective, confident that the market will enthusiastically adopt proven innovations.

Before evidence became important in medicine, agriculture, and technology, products and treatments in each area were disseminated by slick marketing, misleading demonstrations, word of mouth, and tradition. In the nineteenth century, for example, there was already plenty of knowledge in medicine, but neither physicians nor the general public paid consistent attention to it. In the early 1900s, William Halsted, a medical researcher at Johns Hopkins University, spent 30 years trying with limited success to convince physicians to wash their hands before operations and use sterile procedures that had been validated in research going back to the 1860s.

The practice of education today is at much the same pre-scientific point as medicine was a hundred years ago. We have much knowledge in education, and educators do occasionally pay attention to it, as physicians did in 1908. However, there is limited research evaluating specific programs, practices, or materials, and that which does exist is rarely consequential in educators' decisions. As a result important decisions about educational programs are likely to be made based on slick marketing, misleading demonstrations, word of mouth, tradition, and politics. This not only fails to provide the best educational programs to vulnerable children, but it also removes any incentive for developers to create programs and technology that actually work better than current practices. The result is the famous pendulum of educational reform, in which new ideas appear, become widely used, and only then are evaluated. By the time the evaluation evidence is in, the market has already given up on the new idea, and has rushed off to the latest new idea (see Slavin, 1989). A pendulum swing describes innovation in all fields, such as art and fashion, in which taste rather than evidence drives consumer choices. Unfortunately, education is one such field.

Evidence-Based Reform in Education

If education is to make significant progress in the twenty-first century, it must embrace evidence-based reform. There is no other way forward. However, there are great obstacles to be overcome.

For evidence-based reform to prevail, three conditions must exist:

http://dx.doi.org/10.2304/eerj.2008.7.1.124

1. There must be a broad range of proven programs in every area of education, every subject and grade level. Evidence-based policies will not prevail if demanding strong evidence requires educators to use just one or two proven programs, or if no programs have strong evidence.
2. Trusted, impartial, educator-friendly reviews of research must be available, to enable educators and policy makers to know which specific programs and practices have been proven to work in rigorous evaluations.
3. Government agencies must provide incentives to schools to adopt proven programs.

Building the Research Base for Effective Programs

Perhaps the most important requirement for evidence-based reform is the development of a substantial set of replicable programs and practices with strong evidence of effectiveness. Educators and policy makers must have a variety of programs they can choose among with confidence. This means that governments must fund a wide range of research and development projects designed to create innovations capable of significantly improving the outcomes of education at all levels and in all subjects.

The development process might use a series of 'design competitions', in which government sets out what it wants and then funds a variety of entities to develop and then evaluate competing alternatives (Slavin, 1997). For example, a funding agency might ask R&D organizations to develop an approach to teaching algebra to pupils aged 14-16 capable of increasing their performance on national or international measures by at least 25% of a standard deviation. The most promising applicants could be chosen in a competition, and this number might be winnowed down over time if some designs turn out not to be practical. The New American Schools Development Corporation (NASDC), funded by large corporations rather than government, did exactly this to create comprehensive school reform models in the 1990s (Kearns & Anderson, 1996). Initially, 11 design teams were chosen from almost 700 applications. Over time, four of the teams were dropped. Some of the remaining seven, plus a handful of similar programs funded in other ways, ultimately developed strong research bases (Comprehensive School Reform Quality Center [CSRQ], 2006a, b), and at their peak (in 2001) were used in more than 6000 schools in the USA (see Stringfield et al, 1996; Aladjem & Borman, 2006; Slavin, in press). Comprehensive school reform models such as Success for All (Slavin & Madden, 2001), America's Choice (Supovitz et al, 2001), Direct Instruction (Adams & Engelmann, 1996), and the School Development Program (Comer et al, 1996) have been extensively evaluated and found to be effective, and continue to be used in thousands of US schools, even in the absence of government support. Our Success for All program, which focuses on reading in high-poverty elementary schools, has been evaluated in more than 50 experimental–control comparisons, including a national randomized experiment involving 35 schools (Borman et al, 2007) and a nine-year longitudinal follow-up (Borman & Hewes, 2002). It is used in about 1200 schools in the USA and 100 in England.

The design competition process that produced comprehensive school reform models could be applied to any subject and grade level. Imagine design teams working on the design and evaluation of programs capable of accelerating achievement in beginning reading, upper elementary math, algebra, and physical science, programs for second language learners, drop-out prevention, early childhood, and so on. In each case the goal would be to build on the best that currently exists, and to end up with numerous programs, all of which have been proven to increase achievement by at least 25% of a standard deviation. These programs could be highly diverse. Some would involve technology, others not. Some would require extensive training, others less so. Some would challenge current conceptions of curriculum, others might be traditional. All that matters is that they are proven effective and are replicable in many schools.

A key issue in the design competition process is the design of the evaluation. Evaluations should use random assignment to conditions, should use national or state assessments as their outcome measures, and should be large enough (at least 10 schools) to avoid idiosyncrasies due to particular schools. The programs should be implemented under the realistic conditions that will

exist in practice, without extra attention or non-replicable conditions. The evaluations may be done by developers or by third parties, but under close observation by the funding agency.

In the USA, the Institute of Education Sciences (IES) is currently funding a broad array of development and evaluation activities that will ultimately add to the set of proven, replicable models, but due to the anemic funding provided to IES, this process is going too slowly. Governments in many countries could fund substantial research and development of effective programs with a tiny fraction of the money they spend on providing education. With sufficient support, researchers, developers, and entrepreneurs could develop and evaluate programs in every area of pre-kindergarten to secondary education within a period of five to ten years.

Reviewing What Works

Educators and policy makers need to have scientifically valid, fair, and clearly written summaries of the strength of the research evidence showing the effectiveness of education programs. Educators are extremely unlikely to take the time to try to weigh competing evidence from many evaluations. They need information they can rely on in a summative form, like *Consumer Reports* does in the USA and *Which Car?* does in Britain.

The US Department of Education established a website with this objective called the What Works Clearinghouse (WWC; see http://www.whatworksclearinghouse.gov). The WWC provides systematic reviews of research on programs for beginning reading, elementary and middle school mathematics, preschool programs, drop-out prevention, and a few others. Unfortunately, the WWC uses procedures that are strict on random assignment and statistical procedures but pay little attention to use of biased measures or small sample sizes, and as a result its highlighted programs tend to be supported by very small studies (often less than 50 students), very brief studies (often six weeks or less), and studies that use measures of the content taught in the experimental group but not the control group (see Slavin, 2008). With substantial revisions the WWC could still become the pre-eminent source of reviews, but at present it is not useful.

In addition to the What Works Clearinghouse, other websites have sprung up to provide educator-friendly reviews of research on educational programs. The Best Evidence Encyclopedia (http://www.bestevidence.org), from the Center for Data-Driven Reform at Johns Hopkins University, summarizes reviews from all sources, in a *Consumer Reports* format, and contributes its own reviews. It currently has reviews on elementary and secondary mathematics, secondary reading, comprehensive school reform, computer-assisted instruction and other topics.

The Comprehensive School Reform Quality Center (http://www.csrq.org) reviews research on comprehensive school reform models. Other websites, such as the Promising Practices Network (http://www.promisingpractices.net) and Social Programs that Work (http://www.evidencebasedprograms.org) present education and social service programs with the highest levels of evidence, from high-quality randomized experiments. The International Campbell Collaboration (http://www.campbellcollaboration.org) and Britain's EPPI Centre (http://www.EPPI.ioe.ac.uk) also provide scientific reviews in many areas. Although there are controversies and difficulties in program effectiveness reviews (see Slavin, 2008), this enterprise is moving forward rapidly on many fronts, and within a few years it is likely that there will be multiple high-quality, reliable reviews available to educators and policy makers.

Evidence-Based Policies

Ultimately, it is not enough to have many research-proven programs and trusted reviews of research. Education lacks a tradition of looking to evidence for program decisions, and without clear support from government, marketing will always trump evidence.

The US Congress and both the Clinton and Bush administrations have tried to support research-proven practices. The Obey-Porter Comprehensive School Reform program funded schools to adopt 'proven, comprehensive' programs, but most schools funded used programs lacking evidence or even home-grown programs that had never even been piloted before. The Reading Excellence Act promoted research-proven programs, and No Child Left Behind famously

mentioned 'scientifically based research' more than 100 times. Yet this language had little if any impact on practice; even in the Reading First program, which had a particularly strong emphasis on 'scientifically-based research', programs with strong evidence of effectiveness were *less* likely to be adopted in schools receiving Reading First funding than in similar Title I schools that did not receive Reading First funding (Moss et al, 2006). 'Scientifically-based research' turned out to mean 'includes some phonics and is published by a major publisher' (Grunwald, 2006; Manzo, 2006).

The problem with these efforts is that in each case, the legislation contained language supporting research-based practice, but it did not point schools to particular programs with strong evidence. As an analogy, imagine that the FDA just said 'use safe and effective medicines' instead of saying 'Use penicillin. Don't use laetrile'. As a result, publishers and program developers could and did claim research support, and state and federal administrators could and did decide without any rationale what they considered to be 'based on scientifically-based research'. In practice, ambiguous language leaves the issue up to marketing and public relations, not to evidence.

In government policies to support the use of proven programs, it is essential to be clear about which programs have strong evidence of effectiveness. This will become possible in the near future because of the existence of reviews that use consistent standards of evidence, as discussed earlier.

Governments should provide incentives to use programs that have been proven to be effective. One mechanism would be to provide additional rating points in grants for schools or districts applying to use proven programs. Providing additional points instead of requiring use of particular programs allows schools to use any program they think is best, but it clearly expresses a government preference for programs with strong evidence.

Consequences of Evidence-Based Reform

The consequences of evidence-based reform would be profound. If government policies began to favor programs with strong evidence, developers, including publishers, software developers, university researchers, and entrepreneurs of all kinds, would have an incentive to engage in serious development and evaluation efforts. Seeing the immediate impact of research and development, policy makers might provide substantially greater funding for these activities. Developers would have a reason to invest in more effective innovative strategies, knowing that if they turn out to be effective in rigorous evaluations, they will be successful in the marketplace.

Evidence-based reform would finally apply to education the process that led to dramatic developments in medicine, agriculture, and technology in the twentieth century, where every solution that meets evidence standards supersedes less effective products, and a vast R&D enterprise works to improve on the best we have available today.

The winners in this would be millions of children, especially those who are least well served by the current system, the teachers who yearn for more effective tools to help them do their job well, and the whole society, which would come to expect progress in education as confidently as it currently expects progress in other fields. Education research would gain the respect and the resources it has never had. Even the publishing companies that currently rely on marketing would benefit if they embrace innovation, as they would have the resources to do the necessary R&D, just as large drug companies benefit from evidence-based practice in medicine.

Recent developments in research and policy make it possible to finally put education on the road to genuine reform. It remains for policy makers and our profession to take the necessary steps.

Note

[1] Keynote presented at the European Conference on Educational Research, Ghent, Belgium, September 19, 2007.

References

Adams, G.L. & Engelmann, S. (1996) *Research on Direct Instruction: 25 years beyond DISTAR.* Seattle: Educational Achievement Systems.

Robert E. Slavin

Aladjem, D.K. & Borman, K.M. (Eds) (2006) *Examining Comprehensive School Reform.* Washington, DC: Urban Institute Press.

Borman, G. & Hewes, G. (2002*)* Long-Term Effects and Cost-effectiveness of Success for All, *Educational Evaluation and Policy Analysis,* 24(2), 243-266. http://dx.doi.org/10.3102/01623737024004243

Borman, G., Slavin, R.E., Cheung, A., Chamberlain, A.M., Madden, N.A. & Chambers, B. (2007) Final Reading Outcomes of the National Randomized Field Trial of Success for All, *American Educational Research Journal,* 44(3), 701-731. http://dx.doi.org/10.3102/0002831207306743

Comer, J.P., Haynes, N.M., Joyner, E.T. & Ben-Avie, M. (1996) *Rallying the Whole Village: the Comer process for reforming education.* New York: Teachers College Press.

Comprehensive School Reform Quality Center (CSRQ) (2006a) *CSRQ Center Report on Elementary School Comprehensive School Reform Models (revised).* Washington, DC: American Institutes for Research.

Comprehensive School Reform Quality Center (2006b) *CSRQ Center Report on Middle and High School Comprehensive School Reform Models.* Washington, DC: American Institutes for Research.

Grunwald, M. (2006) The Education Issue. http://www.washingtonpost.com (retrieved October 1, 2006).

Kearns, D.T. & Anderson, J.L. (1996) Sharing the Vision: creating new American schools, in S. Stringfield, S. Ross & L. Smith (Eds) *Bold Plans for School Restructuring: the New American Schools designs.* Mahwah: Erlbaum.

Manzo, K. (2006) Federal Review of Reading First Identifies Serious Problems. http://www.edweek.org (retrieved September 22, 2006).

Moss, M., Jacob, R., Boulay, B., Horst, M. & Poulos, J. (2006) *Reading First Implementation Evaluation: interim report.* Cambridge, MA: Abt Associates.

Slavin, R.E. (1989) PET and the Pendulum: faddism in education and how to stop it, *Phi Delta Kappan,* 70, 752-758

Slavin, R.E. (1997) Design Competitions: a proposal for a new federal role in educational research and development, *Educational Researcher,* 26(1), 22-28. http://dx.doi.org/10.2307/1176869

Slavin, R.E. (2008) What Works? Issues in Synthesizing Educational Program Evaluations, *Educational Researcher*, 37(1), 5-14. http://dx.doi.org/10.3102/0013189X08314117

Slavin, R.E. (in press) Comprehensive School Reform, in T. Good (Ed.) *21st Century Education: a reference handbook.* Thousand Oaks: Sage.

Slavin, R.E. & Madden, N.A. (Eds) (2001) *One Million Children: Success for All.* Thousand Oaks: Corwin Press.

Stringfield, S., Ross, S. & Smith, L. (Eds) (1996) *Bold Plans for School Restructuring: the New American Schools designs.* Mahwah: Erlbaum.

Supovitz, J.A., Poglinco, S.M. & Snyder, B.A. (2001) *Moving Mountains: successes and challenges of the America's Choice comprehensive school reform design.* Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

Correspondence: Robert E. Slavin, Institute for Effective Education, University of York, Heslington, York YO10 5DD, United Kingdom (rs553@york.ac.uk).