PEARSON

policy report

Scientifically Based Research: NCLB and Assessment

Sasha Zucker

March 2004

Scientifically Based Research: NCLB and Assessment

A significant aspect of the *No Child Left Behind Act of 2001* (*NCLB*) is the use of the phrase "scientifically based research" well over 100 times throughout the text of the law. The phrase has inspired much discussion concerning its meaning and the corresponding implications for education policy. This report reviews the prevailing interpretations of scientifically based research by both the scientific community and the federal government and examines the importance of these interpretations in policy decisions concerning the funding of educational programs and products. Assessment and testing, which are central to *NCLB's* accountability requirements, serve an important, leading role in scientific education research. Moreover, recent scientific research suggests compelling new applications of educational assessments for improving student achievement.

Scientifically Based Research and NCLB

The debate over the meaning of the phrase "scientifically based research" rose to national prominence with its definition and use in 1990s federal legislation that provided funds to research early reading programs for children (Eisenhart and Towne, 2003). Discussion concerning this topic has increased with the passage into law of *NCLB*, which specifically defines scientifically based research as:

A. research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs; and

B. includes research that:

   i. employs systematic, empirical methods that draw on observation or experiment;

   ii. involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;

iii. relies on measurements or observational methods that provide reliable and valid data across evaluators and observers, across multiple measurements and observations, and across studies by the same or different investigators;

iv. is evaluated using experimental or quasi-experimental designs in which individuals, entities, programs, or activities are assigned to different conditions and with appropriate controls to evaluate the effects of the condition of interest, with a preference for random-assignment experiments, or other designs to the extent that those designs contain within-condition or across-condition controls;

v. ensures that experimental studies are presented in sufficient detail and clarity to allow for replication or, at a minimum, offer the opportunity to build systematically on their findings; and

vi. has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review. (§ 7801 [37])

Federal Interpretation of Scientifically Based Research

Despite extensive discussion, universal agreement about the exact meaning of the definition of scientifically based research in *NCLB* remains elusive (Boghetto, 2003). However, one might convincingly propose that the most relevant interpretation of scientifically based research for those concerned with education research is the interpretation used by federal government agencies in regulations and policy decisions. Throughout *NCLB*, funding for instructional materials and education programs, especially programs from educational service providers, must be justified by evidence from scientifically based research (Beghetto; Eisenhart and Towne, 2003).

Proponents of this approach to research in education have included the National Research Council in their widely cited report *Scientific Research in Education* (Shavelson and Towne, 2002). The report asserts that education is a social science that can be investigated using randomized experiments of the design successfully used in other social sciences (such as psychology and economics) as well as in the "hard" sciences (such as physics and biochemistry). This view is further supported through research that compellingly suggests that scientific studies of education can be of comparable rigor to research in the "hard" sciences (Marzano, Pickering, Pollock, 2001, p. 4).

Dr. Grover J. Whitehurst, Director of the U.S. Department of Education's Institute of Education Sciences, echoes this view by drawing a parallel between the progress of research in education and progress in other applied sciences such as medicine and

agriculture. He proposes that education is an applied science that can be improved in the same way (Whitehurst, 2002a). Hence, *NCLB's* emphasis on scientifically based research is intended to change the approach that researchers take in determining "what works" (Mageau, 2004) in education to reflect the U.S. Food and Drug Administration's requirement that the effectiveness of medications and therapies be evaluated using randomized clinical trials (Olson and Viadero, 2002).

To clarify what research is considered scientifically based, the U.S. Department of Education recently proposed that, when considering funding for educational practices or programs, priority be given to programs supported by research that uses "an experimental design under which participants—e.g., students, teachers, classrooms, or schools—are randomly assigned to participate in the project activities being evaluated or to a control group that does not participate in the project activities being evaluated" (U.S. Department of Education, 2003a, p. 62446). This proposed interpretation raises the expected standard for education research to the level of research used in the other social sciences as well as in the "hard" sciences (Shavelson and Towne, 2002). The random assignment of research participants to experimental and control groups is frequently referred to as the "gold standard" of research (Slavin, 2003; Trybus, 2004).

Characteristics of Experimental Research in Education

The assertion that education can be investigated using randomized experiments is not uncontroversial; the literature of this debate starts as early as the nineteenth century (Boghetto, 2003), and proponents of experimental research in education acknowledge the challenges presented by the field. The National Research Council (Shavelson and Towne, 2002) notes the particular aspects of education that researchers must consider, including the values and politics of the various stakeholders (such as children, parents, teachers, and policy makers); human volition (for example, a family's decision to move to a different geographic area before a study's completion); variability in education programs; the wide variety in the organization of schools; and the diversity of the many individuals involved in education (geographical, historical, social, ethnic, linguistic, economic, and cultural).

Adding to the complexity of education research is the very character of education research itself. Aspects of education research noted by the National Research Council (Shavelson and Towne, 2002) include the multiple-disciplinary perspectives on education (contributions to education research frequently come from other fields, such as sociology, psychology, economics, and even from the "hard" sciences such as physics and biology); ethical considerations (the protection of human participants in research, especially children); and relationships and partnerships between researchers and those engaged in professional education practice (teachers, administrators, curriculum developers, and so on). As with the other social sciences, the most evident characteristics of education research are the inherent complexities and lack of control that come with studying humans. It has been astutely observed that "social scientists

lack the high degree of control over their subjects that is typical in the 'hard' sciences—for example, gaggles of molecules are better behaved than a classroom of third-graders" (Shavelson and Towne, p. 48).

Scientific Research Designs

This complexity has led many to question whether randomized experiments are even possible in education research (Olson and Viadero, 2002; Olson, 2004). Moreover, some assert that giving priority to scientifically based research will result in a narrow scope that ignores existing, useful research that does not meet *NCLB's* preference for randomized experimental design (Olson and Viadero; Mendez, 2004). However, proponents of scientifically based education research recognize that research methods can take a variety of forms to account for the variables and ethical concerns outlined above. The U.S. Department of Education (2003a) lists other research methods that can be used when random assignment is not feasible, such as *quasi-experimental* research. This design achieves an effect close to a truly randomized experiment by matching the randomly assigned participants to subjects who are not participants in the study, but who can serve as a control group by virtue of similar characteristics.

Others propose that certain research designs common in education research, such as case studies, may qualify as scientific by adhering to certain general scientific principles (Shavelson and Towne, 2002; The Charles Dana Center, 2003). In their view, research can be accepted as scientific if it poses significant questions that can be investigated empirically; links research to relevant theory; uses methods that permit direct investigation; provides a coherent and explicit chain of reasoning; replicates and generalizes across studies; and encourages professional scrutiny and technique. As *NCLB's* definition cites only a preference for experimental design, other valid research designs that adhere to these general scientific principles may qualify as scientifically based research.

Evidence-based Education

The definition of scientifically based research in *NCLB* does not appear to mean that it "is the 'only' or best research that should be funded" (The Charles A. Dana Center, 2003, p. 3) by the federal government. Whitehurst (2002a) makes a distinction between scientifically based research in education and evidence-based education (see Figure 1). This inclusive approach to improving education integrates professional wisdom (the judgment that individuals acquire through experience) with empirical evidence (of which scientifically based research is only one, albeit an essential, source) to determine "what works" (Mageau, 2004). While *NCLB* emphasizes stricter scientific designs for future research, existing research still serves a purpose within the framework of evidence-based education. *NCLB*'s emphasis on scientifically based research with a preference for randomized experiments is clearly not intended to discount other data sources from contributing to the improvement of education.

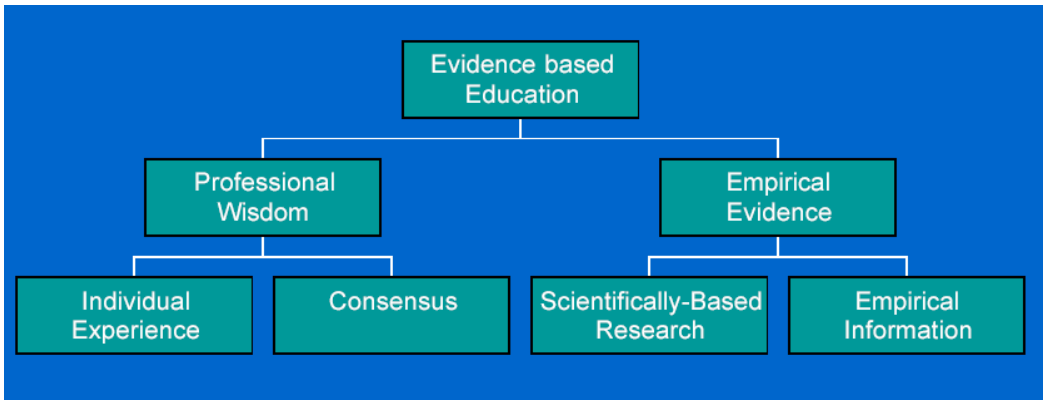

Figure 1. Components of evidence-based education (Whitehurst, 2002a).

Increasing the Amount of Education Research Available

A final challenge to *NCLB*'s strict standards for scientifically based research is that, except for the research conducted by test publishers such as Pearson Education, Inc. (Pearson), not enough research meeting such standards is currently available. However, one can argue that this lack of evidence from randomized experiments is precisely why a mandate for scientifically based research is necessary (Slavin, 2004). Moreover, the requirement for evidence from scientifically based research as a justification for federal funding of products and programs will motivate education companies to take responsibility for investing in clinical trials and evaluations (Olson and Viadero, 2002). By requiring scientifically based research, *NCLB* provides education leaders with a way to ensure that federal funding is being spent on products and programs supported by objective evidence—products and programs that work.

Scientifically Based Research and Assessment

Educational assessments and scientific research concerning education are closely related. Past scientific research has firmly established the utility of educational assessments; scientific research depends on educational assessments as a source of raw data; and recent scientific research indicates that well designed assessments may improve student achievement as well as measure it.

History of Education Research and Assessment

Over the past 150 years, much of the scientific research in education has concerned the development and use of tests and assessments. Through a series of scientific studies following the strict standards discussed above, researchers in education, psychology, and psychometrics have developed rigorous, objective methods for statistically measuring an assessment's reliability (the consistency of the scores it produces) and validity (the degree to which scores accurately measure the traits that

they are intended to measure). These measurements of reliability and validity are widely used today by leading test publishers, such as Pearson, as evidence of an assessment's quality. Moreover, scientific research in assessment has led to mathematical models, such as Item Response Theory (IRT), that allow detailed analyses of test results. IRT allows for the analyses of results at the level of individual items and has a variety of sophisticated applications, such as reporting scores on a common vertical scale—as with Pearson's Stanford Scale—and the creation of adaptive tests to be taken on computers (Shavelson and Towne, 2002).

Although *NCLB* does not specifically require that assessments be supported by scientifically based research, the measures that support an assessment's validity, reliability, and other technical qualities are well established by scientifically based research. To meet the accountability mandate of *NCLB*, educational agencies are required to use assessments that are "valid and reliable, and consistent with relevant, nationally recognized professional and technical standards" and also "of adequate technical quality for each purpose required under [the] Act" (§ 6311). Hence, it increasingly important for administrators to understand the scientific evidence used to support assessments (Mann and Shakeshaft, 2003).

Assessments Providing Data in Scientific Education Research

Assessments also play an essential role in scientific education research. The results of academic achievement tests, developed using the rigorous methods described above, are the main source of data for evaluating the outcomes of experimental education practices (Marzano, Pickering, and Pollock, 2001). Hence, it is crucial that the assessments used in research studies be of the highest quality. As an important criterion for the evaluation of a research study, the U.S. Department of Education recommends the collection of data on student achievement using an assessment "whose ability to accurately measure true skill levels is well established" and cites Pearson's *Stanford Achievement Test* Series as an assessment that meets this requirement (2003b, p. 6).

Research Indicating Educational Benefits of Achievement Testing

NCLB also allows for funding in support of the development of assessments to improve student achievement and meet accountability requirements (§ 7301). For example, scientific research has revealed further benefits from educational assessments beyond the mere measurement of student achievement. Findings from several scientific research studies (Marzano et al., 2001) suggest that feedback from test results can significantly contribute to improving student achievement, especially if the feedback is timely, corrective (provides an explanation for inaccurate responses to items), and specific to a criterion (for example, from academic content standards). Additionally, students can use test results to track their own progress, thereby using self-evaluation to improve their achievement. This research suggests that a

norm-referenced assessment—such as the *Stanford Achievement Test* Series, Tenth Edition (Stanford 10)—that is augmented to align with state content and performance standards can serve the twofold purpose of measuring student achievement to meet *NCLB* accountability requirements while also helping to improve student achievement scores. Additional benchmark assessments administered between high-stakes testing periods may serve as an invaluable opportunity to assist students in raising their levels of achievement.

Conclusion

Scientifically based research, as defined and mandated by *NCLB*, has many implications for educational policy, especially in regard to the development and application of educational assessments. Pearson leads the test publishing industry in the use of rigorous, time-tested scientific research studies to support the reliability and validity of its assessment products. Pearson's Research Advisory Board, which includes academic and professional experts in the fields of applied research and psychometrics, provides assistance and expertise in the latest assessment and education research methodologies. Pearson's scientific research provides evidence that the innovative features of Stanford 10—such as the full-color testing materials, untimed subtests, and universal design—enhance students' ability to demonstrate accurately what they know and are able to do.

By requiring evidence from scientifically based research to justify funding for educational programs and activities, *NCLB* provides decision makers with a valuable tool for ensuring that high-quality products are obtained from education service providers. Holding education research to high scientific standards will contribute to a better understanding of what works in education, improving the educational opportunities of future generations.

References

Beghetto, R. (2003, April). Scientifically-based research. *ERIC Digest, 167*. Retrieved from http://eric.uoregon.edu/publications/digests/digest167.html on February 19, 2004.

The Charles A. Dana Center. (2003, April). Making sense of research for improving education. *Dana Center Issue Brief, 5.*

Eisenhart, M., & Towne L. (2003, October). Contestation and change in national policy on "scientifically based" education research. *Educational Researcher 32(7),* pp. 31–38.

Mageau, T. (2004, January). Determining "what works." *T H E Journal (Technological Horizons in Education), 31(6),* p. 32.

Mann, D., & Shakeshaft, C. (2003, January). In God we trust; All others bring data. *School Business Affairs.* Retrieved from http://asbointl.org/asbo/files/ccPageContentdocfilename001016705546SBA_Jan_03_InGodWeTrust_article.pdf on February 24, 2004.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Mendez, T. (2004, January 29). Reading choices narrow for schools with federal aid. *The Christian Science Monitor,* p. 12.

No Child Left Behind Act of 2001, 20 U.S.C. § 6301 *et seq.* (2001).

Olson, D. (2004, January). The triumph of hope over experience in the search for "what works": a response to Slavin. *Educational Researcher 33(1),* pp. 24-26.

Olson, L., & Viadero, D. (2002, January 30). Law mandates scientific base for research. *Education Week,* p. 1.

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education.* Washington, DC: National Academy Press.

Slavin, R. E. (2003, February). A reader's guide to scientifically based-research. *Educational Leadership, 60(5),* pp. 12–16.

Trybus, M. E. (2004). The challenge and hope of scientifically based research. *Viewpoints, 11*, pp. 3–21.

U.S. Department of Education. (2003a, November 4). Scientifically based evaluation methods. *Federal Register 68(213),* pp. 62445–62447.

U.S. Department of Education. (2003b, December). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide.* Washington, DC: Author.

Whitehurst, G. J. (2002a). *Evidence-based education (EBE).* Washington, DC: U.S. Department of Education. Retrieved from http://www.ed.gov/admins/tchrqual/evidence/whitehurst.ppt on Feburary 23, 2004.

Whitehurst, G. J. (2002b, June). *Statement of Grover J. Whitehurst, Assistant Secretary for Research and Improvement, before the Senate Committee on Health, Education, Labor, and Pensions.* Washington, DC: U.S. Department of Education. Retrieved from http://www.ed.gov/news/speeches/2002/06/06252002.html on February 23, 2004.

Additional copies of this and related documents are available from:
Pearson Education, Inc.
19500 Bulverde Road
San Antonio, TX 78259
1-800-211-8378
1-877-576-1816 (fax)
http://www.pearsonassess.com