1. Introduction

Among educational leaders and policymakers there has been increasing concern regarding the need for scientifically based evidence on which to base funding decisions for specific educational programs and practices. This concern is fundamentally about having better evidence for making decisions about what programs and practices do or do not work. The need for such evidence leads to causal questions, such as whether particular programs and practices improve student academic achievement, social development, and educational attainment. Issues of causality are not new to the academy or public debate and have a rich history in disciplines such as psychology and philosophy and in specialized fields of education research. Nonetheless, among researchers there is a lack of clarity regarding which designs, methods, and analytic approaches are most appropriate for making causal inferences. This report is intended to help researchers, educators, and policymakers understand causal estimation by describing the logic of causal inference and reviewing designs and methods that allow researchers to draw causal inferences about the effectiveness of educational interventions.

Recently, questions of causality have been at the forefront of educational debates and discussions, in part because of dissatisfaction with the quality of education research and recent federal initiatives designed to promote the accumulation of scientific evidence in education that rely on randomized controlled trials (RCTs). A common concern expressed by those deeply engaged with the educational enterprise, as well as those outside education, revolves around the design of and methods used in education research, which many claim have resulted in fragmented and often unreliable findings (Kaestle, 1993; Levin & O'Donnell, 1999; Sroufe, 1997). Pointing to lack of replication, inappropriate designs for assessing causal effects, and crude analytic procedures, some researchers have argued that it is difficult to accumulate a knowledge base that has value for practice or future study (Cook, 2002; Lagemann, 1999, 2000; Shavelson & Berliner, 1988; Weiss, 1999). Education researchers have long struggled with the need to balance "pure" research with the discovery of "what works" and to evaluate the strengths and weaknesses of various methodologies for addressing particular research questions. Several new national initiatives have brought these issues of methodological rigor to the forefront.

First, The No Child Left Behind Act of 2001 (NCLB) provided a specific definition of scientifically based research and set aside funding for education research studies consistent with that definition. Second, funded research programs in the Institute of Education Sciences (IES), the National Science Foundation (NSF), and the National Institute of Child Health and Human Development (NICHD) have increased calls for intervention studies that provide clear evidence of student learning (for details of this history see Eisenhart & Towne, 2003). Third, projects undertaken by the National Academy of Sciences' National Research Council (NRC) have produced a series of reports focused on improving the quality of education research (see, e.g., NRC, 2002, 2004a).

Evidence-based research is one of the four pillars of NCLB, which places special emphasis on determining, through rigorous scientific study, which educational programs and practices are effective. This concept was reinforced by the Education Sciences Reform Act of 2002, which replaced the Office of Educational Research and Improvement with the newly created IES. The goal of IES is "the transformation of education into an evidence-based field in which decision makers routinely seek out the best available research and data before adopting programs or practices that will affect significant numbers of students" (IES, http://www.ed.gov/about/offices/list/ies/index.html?src=oc). This legislation states that by conducting scientifically based research studies that apply rigorous, systematic, and objective methodology to obtain reliable and valid knowledge, it is possible to identify educational practices and activities that result in improved student learning.

These goals were reiterated in a statement from the Secretary of Education in the *Federal Register* (2005), which noted that "random assignment and quasi-experimental designs [are considered] to be the most rigorous methods to address the question of project effectiveness" (p. 3586). While these designs have particular importance for programs authorized by NCLB and IES, they are also being established as a priority for all U.S. Department of Education programs.

The press for randomized controlled trials is illustrated by the What Works Clearinghouse, a federally funded organization that reviews results of randomized trials that have demonstrated beneficial causal relationships between educational interventions and student outcomes, such as improving early reading comprehension and reducing high school dropout rates. Guided by a technical advisory panel, the What Works Clearinghouse has established quality standards to review available research, placing a high priority on randomized field trials, which are seen as being "among the most appropriate research designs for identifying the impact or effect of an

education program or practice" (What Works Clearinghouse, http://www.w-w-c.org). Acknowledging that randomized field trials are not feasible in certain situations or for some research questions, the What Works Clearinghouse also advocates the use of quasi-experiments, that is, comparative studies that carefully attempt to isolate the effect of an intervention through means other than randomization.

IES and other federal agencies have made the use of randomized controlled trials a research priority, and several large-scale studies are currently under way. In addition, NSF, IES, and NICHD have collaborated in sponsoring the Interagency Education Research Initiative (IERI), a program that is explicitly designed to bring promising educational interventions to scale. The program promotes research that identifies promising interventions through rigorous randomized trials; when there are justifiable results, it in turn supports the replication of interventions with other groups of participants. To date, the IERI has approximately 108 projects either in the developmental testing phase of an intervention or in the process of bringing an intervention to scale (McDonald, Keesler, Kauffman, & Schneider, 2006; Schneider & McDonald, 2007a, 2007b).

Another program designed to review evidence-based research in education has been undertaken by the NRC, the operating arm of the National Academies. In 2000, NRC established a committee to "review and synthesize recent literature on the science and practice of scientific educational research and consider how to support high-quality science in a federal education research agency" (NRC, 2002, p. 22). The committee, composed of scholars with diverse disciplinary affiliations and varied methodological expertise, carried out several activities aimed at identifying what constitutes scientific research and how the principles of science can be translated into education research. The committee published the monograph *Scientific Research in Education* (NRC, 2002), which provides

an articulation of what constitutes high-quality scientific inquiry in education. A follow-up committee produced a second report, *Advancing Scientific Research in Education* (NRC, 2004a), which recommends ways to promote evidence-based research in education.

The two committees concluded that, when financially, logistically, and ethically feasible, the randomized field trial is the best design for making causal inferences about the effectiveness of educational programs and practices. However, they also emphasized that random assignment is not always warranted, feasible, or ethical and recommend the use of quasi-experiments and statistical modeling in these instances.[4]

The NRC maintained that one way to shape the understanding of what constitutes high-quality scientific research is to create and sustain a set of norms and common discourse in the educational community regardless of methodological differences (see also Feuer, Towne, & Shavelson, 2002). To this end, the first committee outlined a set of six guiding principles that it concluded should underlie all scientific inquiry. These principles formed the core of *Scientific Research in Education* (NRC, 2002), which continues to have a significant influence both nationally and internationally among scientists in education and other fields. (Both *Educational Researcher* and *Teachers College Record* have published theme issues on it ["Scientific research," 2002; "Scientific research," 2005]; also see Giangreco & Taylor, 2003; Horner, Carr, Halle, McGee, Odom, & Wolery, 2005; Kamil, 2004; Lagemann, 2005; Levin, 2003; Mayer, 2003; Rover, 2005; Stoiber, 2002; Thompson, Diamond, McWilliam, Snyder, & Snyder, 2005.)[5]

Consensus on these guidelines has not been reached within the education research community. There are a number of researchers who are at odds with the philosophical and methodological value placed on scientific principles as a basis for understanding the implications and consequences of educational reforms for students and their teachers (Bloch, 2004;

Fish, 2003; Gee, 2005; Lather, 2004; Moss, 2005; Popkewitz, 2004; Spooner & Browder, 2003; Willinsky, 2005). Despite these concerns, the report has become a catalyst for discussion and action in schools and colleges of education and among scholars and policymakers in the NRC and the federal government, primarily because of its clear message regarding what should and should not be considered scientific evidence.

The NRC reports represent instances where diverse independent committees have stated that designs and methods for conducting education research are not equally suitable for addressing particular questions. The committees concluded that research designs should be carefully selected and implemented to best address the question being studied. Not only do certain problems require different designs, but more important, "some methods are better than others for particular purposes, and scientific inferences are constrained by the type of design employed" (NRC, 2002, p. 98).

Recognizing the need to develop criteria for determining which designs and methods are most appropriate for addressing causal questions, NSF has undertaken a systematic review of its education research portfolio with the goal of developing funding guidelines for future solicitations. To assist the agency in this effort, NSF enlisted the support of the AERA Grants Board to evaluate various research designs and their appropriateness for making causal inferences. Specifically, NSF charged the Grants Board with

> (1) defining causal effects, highlighting the strengths and weaknesses of various study designs intended to examine such effects, and describing analytic methods for estimating effects with different types of study designs; (2) reviewing and selecting NSF-supported studies that demonstrate "scientifically-based research" in which appropriate causal inferences are made; and (3) identifying criteria for designing future studies addressing causal effects. (Memorandum, October 6, 2003)

This charge is consistent with the concern of the AERA Grants Board that education researchers become aware of the strengths and weaknesses of various designs and methods for addressing causal questions. In responding to NSF's charge, the present report describes and exemplifies the role of causal inference in providing evidence on the effectiveness of educational programs and practices. Section 2 begins by defining cause and effect and then reviews the logic of causal inference, presents a formal specification of the causal inference model used in randomized controlled experiments, and provides criteria for making such inferences. Section 3 describes alternative designs that have been developed to approximate randomized experiments. Section 4 summarizes four NSF-supported studies that vary in design and methods of analysis. The strengths and limitations of these designs for making causal inferences are reviewed. The report concludes with recommendations intended to assist NSF and other funding agencies in the review and development of their education research portfolios. These recommendations should also prove useful to researchers and policymakers in developing criteria and guidelines for conducting rigorous scientific research on the effectiveness of educational programs and practices. These types of studies may be especially valuable for policymakers in establishing and assessing funding priorities.

This report is not intended to be a definitive guide for researchers interested in conducting studies based on experimental and observational data. Its purpose is to explain the value of quasi-experimental techniques that can be used to approximate randomized experiments. The report does not address many of the nuances of experimental and quasi-experimental designs. The goal is to describe the logic of causal inference for researchers and policymakers who are not necessarily trained in experimental and quasi-experimental designs and statistical techniques.

Chapter 1 Notes

4 There are other examples of increased interest in the use of multiple methods for making causal inferences. Groups such as the American Psychological Association, AERA, and federal agencies such as NSF have held meetings within the last few years to address how multiple methods can inform causal inference. Also see Raudenbush (2005) on multimethod research and causal analysis.

5 See Schneider, McDonald, Brown, Schalliol, Makela, Yamaguchi, et al. (2006), for a summary of varying perspectives on *Scientific Research in Education* by authors who cited the report.